

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:25:21 ; Search time 28.82 Seconds  
(without alignments)  
3409.477 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 568  
Sequence: 1 MGFLHFLTGILCSLPISL.....TEGDAKNPPVLAGLLIKIVE 568

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	100.0	568	16	Q929G0 chlamydia p
2	12	2.1	566	16	Q84354 chlamydia t
3	10	1.8	566	16	Q9PK44 chlamydia m
4	9	1.6	240	2	Q9F3N4 streptomyces
5	9	1.6	408	2	Q9K529 pseudomonas
6	9	1.6	408	2	Q9AEY2 pseudomonas
7	9	1.6	425	16	Q9KJY9 rhizobium m
8	9	1.6	445	2	Q81411 rhodobacter
9	9	1.6	502	10	Q82752 arabidopsis
10	8	1.4	134	2	Q49007 mycoplasma
11	8	1.4	137	16	Q9PKF9 chlamydia m
12	8	1.4	173	17	Q970S5 sulfolobus
13	8	1.4	197	16	Q97P95 streptococcus
14	8	1.4	206	2	Q54770 streptococcus
15	8	1.4	323	5	Q9B045 leishmania
16	8	1.4	337	11	Q91WE1 mus musculus

# ALIGNMENTS

17	8	1.4	357	2	Q68120 rhodobacter
18	8	1.4	411	2	Q51747 pseudomonas
19	8	1.4	411	2	Q51851 pseudomonas
20	8	1.4	411	2	Q9KAS1 pseudomonas
21	8	1.4	411	2	P95569 pseudomonas
22	8	1.4	412	1	Q9HH00 thermococcus
23	8	1.4	414	2	Q56996 zymomonas m
24	8	1.4	416	17	Q58762 pyrococcus
25	8	1.4	434	16	Q92C76 listeria in
26	8	1.4	435	16	Q9WZJ3 thermotoga
27	8	1.4	435	16	Q9KAZ4 bacillus ha
28	8	1.4	435	16	Q9JUM0 staphylococcus
29	8	1.4	447	16	Q9CG79 lactococcus
30	8	1.4	448	16	Q99ZL9 streptococcus
31	8	1.4	456	16	Q97R84 streptococcus
32	8	1.4	477	10	Q9SEF45 arabidopsis
33	8	1.4	483	5	Q9XV67 caenorhabditis
34	8	1.4	509	2	Q93RV5 streptomyces
35	8	1.4	558	10	Q93YN2 arabidopsis
36	8	1.4	566	1	P77943 sulfolobus
37	8	1.4	570	10	Q948K1 citrus unsh
38	8	1.4	570	10	Q93Y81 citrus para
39	8	1.4	587	16	Q9CGY7 lactococcus
40	8	1.4	596	5	Q950F9 helicoverpa
41	8	1.4	629	10	Q9FX95 arabidopsis
42	8	1.4	629	10	Q91774 arabidopsis
43	8	1.4	667	16	P71749 mycobacterium
44	8	1.4	998	5	Q44800 caenorhabditis
45	7	1.2	47	10	Q9M3R9 arabidopsis

# RESULT 1

Q929G0	PRELIMINARY;	PRT;	568 AA.
AC Q929G0;			
DT 01-MAY-1999 (Trembl)	10, Created		
DT 01-MAY-1999 (Trembl)	10, Last sequence update		
DT 01-DEC-2001 (Trembl)	19, Last annotation update		
DE PUTATIVE OMP PROTEIN.			
GN CPN0021 OR CPJ0021 OR CP0755.			
OS Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX NCBI_TaxID=83558;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=CM1029;			
RX MEDLINE=99206606; PubMed=10192388;			
RA Kilmann S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA Olinger L., Grimwood J., Davits R.W., Stephens R.S.,			
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."			
RL Nat. Genet. 21:385-389(1999).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC STRAIN=AR39;			
RX MEDLINE=20150255; PubMed=10684935;			
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,			
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA Eisen J., Fraser C.M.			
RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia			
RT pneumoniae AR39."			
RL Nucleic Acids Res. 28:1397-1406(2000).			
RN [3]			
RP SEQUENCE FROM N.A.			
RC STRAIN=J138.			
RX MEDLINE=20330349; PubMed=10871362;			
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,			
RA Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.,			
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138			

RT from Japan and CWL029 from USA.  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL: AE001587; AAD18174.1; -  
 DR EMBL: AE002234; AAF38558.1; -  
 DR EMBL: AP002545; BAA98233.1; -  
 DR TIGR: CP0755; -  
 DR InterPro: IPR004155; HEAT\_PBS.  
 DR Pfam: PF03130; HEAT\_PBS; 4.  
 KW Complete proteome.  
 SQ SEQUENCE 568 AA; 63553 MW; 3358FCB5BC482E80 CRC64;

Query Match 100.0%; Score 568; DB 16; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLFHLTFLGGLLGLSPISLVAKFPESVGHKILYSTOSTOQALATYLEALDAYGDHDF 60  
 DB 1 MGLFHLTFLGGLLGLSPISLVAKFPESVGHKILYSTOSTOQALATYLEALDAYGDHDF 60  
 QY 61 VLKRGEDYLKOSTHSSDPQTRKSTIIAGLAGSSEALDVLISQAMETADPLQQLLVLSAV 120  
 DB 61 VLKRGEDYLKOSTHSSDPQTRKSTIIAGLAGSSEALDVLISQAMETADPLQQLLVLSAV 120  
 QY 121 SGHLGKTSDDLFLKALASPYVIRLEAAYRLANKTKNTKVIDHLSFHKLPDEIOCLSA 180  
 DB 121 SGHLGKTSDDLFLKALASPYVIRLEAAYRLANKTKNTKVIDHLSFHKLPDEIOCLSA 180  
 QY 181 IFLRLETESDAYIRDLAARSAIRATALQIGEVQOKREPLRLNLTLSASPODOBAI 240  
 DB 181 IFLRLETESDAYIRDLAARSAIRATALQIGEVQOKREPLRLNLTLSASPODOBAI 240  
 QY 241 LYALGKLDGOSYNNKKOLKOPVDVTLAAQAALIALGKEEDALPVIKKQALEERPAL 300  
 DB 241 LYALGKLDGOSYNNKKOLKOPVDVTLAAQAALIALGKEEDALPVIKKQALEERPAL 300  
 QY 301 YALRHLPSEIGIPALPFLTKNSEAKLNAVALLEIGCPTKLEITYERLVOPHNE 360  
 DB 301 YALRHLPSEIGIPALPFLTKNSEAKLNAVALLEIGCPTKLEITYERLVOPHNE 360  
 QY 361 TLALSFSKGRTLQNMKRVNIIVPODERERLSTTRGLEBOILTFRLRKAVLPCTY 420  
 DB 361 TLALSFSKGRTLQNMKRVNIIVPODERERLSTTRGLEBOILTFRLRKAVLPCTY 420  
 QY 421 KLASQRTQATTAISFISHTSHOEADLFLQAAKLPGEPIIRAYADIAIYNLTKDPERK 480  
 DB 421 KLASQRTQATTAISFISHTSHOEADLFLQAAKLPGEPIIRAYADIAIYNLTKDPERK 480  
 QY 481 RSLHDYAKKLIOETLLPVDTEENORPHSPMYLRYQVTPESRTKIMDLTETLAKSSED 540  
 DB 481 RSLHDYAKKLIOETLLPVDTEENORPHSPMYLRYQVTPESRTKIMDLTETLAKSSED 540  
 QY 541 IRLLIQMTGEDAKNFPYLAGLTIKIVE 568  
 DB 541 IRLLIQMTGEDAKNFPYLAGLTIKIVE 568

RESULT 2  
 ID 084354 PRELIMINARY; PRT; 566 AA.  
 AC 084354;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOHETICAL 63.5 KDA PROTEIN.  
 GN CT350.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID-813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/UM-3/CX;  
 RX MEDLINE=99000809; PubMed=9784136;

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatuzov R.L., Zhao Q., Koonin E.V.,  
 RA Davys R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis."  
 RL Science 282:754-759(1998).  
 DR EMBL: AE001308; AAC67945.1; -  
 DR InterPro: IPR004155; HEAT\_PBS.  
 DR Pfam: PF03130; HEAT\_PBS; 5.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 566 AA; 63507 MW; 92E5D4F190B23D56 CRC64;

Query Match 2.1%; Score 12; DB 16; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 0.0061;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 PYIRLEAAYRLA 152  
 DB 139 PYIRLEAAYRLA 150

RESULT 3  
 ID 09PK44 PRELIMINARY; PRT; 566 AA.  
 AC 09PK44;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOHETICAL PROTEIN TC0629.  
 GN TC0629.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MORN / NIGG;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Baas S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwyn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia  
 RT pneumoniae AR39."  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL: AE002331; AAF39458.1; -  
 DR TIGR: TC0629; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 566 AA; 63143 MW; F9C7B8E7B4E1C6C2 CRC64;

Query Match 1.8%; Score 10; DB 16; Length 566;  
 Best Local Similarity 100.0%; Pred. No. 0.61;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 IRLFAAYRLA 152  
 DB 141 IRLFAAYRLA 150

RESULT 4  
 ID 09F3N4 PRELIMINARY; PRT; 240 AA.  
 AC 09F3N4;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE HYPOHETICAL FUSION PROTEIN.  
 GN SC10F4.26C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;  
 OC Actinomycetales; Streptomycinae; Streptomyces.

OX NCBI\_TaxID=1902;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Seeger K.V., Harris D.;  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA Cereno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,  
 RA Kinsahl H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL450350; CAC16986.1; -;  
 DR InterPro: IPR004360; Gly\_bleo\_diox.  
 DR Pfam: PF00303; Glyoxalase; 1.  
 SQ SEQUENCE 240 AA; 2515 MW; EC8BA5C65D1D3DE6 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 2.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 LTSASPQDQ 237  
 DB 69 LTSASPQDQ 77

RESULT 5  
 ID Q9K529 PRELIMINARY; PRT; 408 AA.  
 AC Q9K529;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE FERREDOXIN REDUCTASE.  
 GN BPH44.  
 OS Pseudomonas sp. B4.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=59381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B4;  
 RA Rodarie D., Willison J., Jounneau Y.;  
 RT "Cloning, purification and molecular characterization of the biphenyl  
 RT dioxygenase complex from Pseudomonas sp. strain B4.";  
 RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- COFACTOR: FAD (BY SIMILARITY).  
 DR EMBL: AJ251217; CAB93969.1; -;  
 DR InterPro: IPR000759; Adnrx\_reductase.  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR InterPro: IPR001100; Pyridine\_redox\_2.  
 DR InterPro: IPR003042; Rng\_mnoxygenase.  
 DR Pfam: PF00070; pyr\_redox; 1.  
 DR PRINTS: PR00419; ADXRDTASE.  
 DR PRINTS: PR00368; FADPNR.  
 DR PRINTS: PR00411; PNDRDTASEI.  
 DR PRINTS: PR00469; PNDRDTASEII.  
 DR PRINTS: PR00420; RNCMNKGXNASE.  
 DR FAD: Flavoprotein; Oxidoreductase.  
 KW SEQUENCE 408 AA; 42911 MW; 81F3BB10642D06B8 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAGS 94  
 DB 7 IIGAGLAGS 15

RESULT 6  
 ID Q9AEY2 PRELIMINARY; PRT; 408 AA.  
 AC Q9AEY2;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE FERREDOXIN REDUCTASE.  
 GN BPHG.  
 OS Pseudomonas sp. Cam-1.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=85672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CAM-1;  
 RX MEDLINE=21268664; PubMed=11375179;  
 RA Master E.R., Mohn W.W.;  
 RT "Induction of bpha, Encoding Biphenyl Dioxygenase, in Two  
 RT Polychlorinated Biphenyl-Degrading Bacteria, Psychrotolerant  
 RT Pseudomonas Strain Cam-1 and Mesophilic Burkholderia Strain LB400.";  
 RL Appl. Environ. Microbiol. 67:2669-2676(2001).  
 CC -1- COFACTOR: FAD (BY SIMILARITY).  
 DR EMBL: AY027651; AKK14785.1; -;  
 DR InterPro: IPR000759; Adnrx\_reductase.  
 DR InterPro: IPR001327; FAD\_pyr\_redox.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR InterPro: IPR001100; Pyridine\_redox\_2.  
 DR InterPro: IPR003042; Rng\_mnoxygenase.  
 DR Pfam: PF00070; pyr\_redox; 1.  
 DR PRINTS: PR00419; ADXRDTASE.  
 DR PRINTS: PR00368; FADPNR.  
 DR PRINTS: PR00411; PNDRDTASEI.  
 DR PRINTS: PR00469; PNDRDTASEII.  
 DR PRINTS: PR00420; RNCMNKGXNASE.  
 DR FAD: Flavoprotein; Oxidoreductase.  
 KW SEQUENCE 408 AA; 42982 MW; 975E2009AA387979 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAGS 94  
 DB 7 IIGAGLAGS 15

RESULT 7  
 ID Q9KJY9 PRELIMINARY; PRT; 425 AA.  
 AC Q9KJY9;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 2-OCTAPRENTYL-6-METHOXYPHENOL HYDROXYLASE (PUTATIVE OXIDOREDUCTASE  
 DE PROTEIN).  
 GN UBH OR SMC00316.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=20317087; PubMed=10858449;  
 RA Schlenker C., de Rudder K.E.E., Kohrs V., Lopez-Lara I.M., Geiger O.,  
 RT "Cloning and characterization of the gene for phosphatidylcholine  
 RT synthase."  
 RL J. Biol. Chem. 275:18919-18925(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21368234; PubMed=11474104;  
 RA Galibert F., Finan T.M., Long S.R., Puchler A., Abola P., Ampe F.,  
 RA Barloy-Hubler F., Barnett M.J., Becker A., Bolstad P., Bothe G.,  
 RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,  
 RA Cowie A., Davis R.W., Dreno S., Federspiel N.A., Fisher R.F.,  
 RA Gloux S., Godtke T., Goffeau A., Golding B., Guzy J., Gurjal M.,  
 RA Hernandez-Lucas I., Hong A., Hulzar L., Hyman R.W., Jones T., Kahn D.,  
 RA Kahn M.L., Kallman S., Keating D.H., Kiss E., Komp C., Lelaure V.,  
 RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,  
 RA Ramsperger U., Surzycki R., Thebaud P., Vandenbol M., Yeh K.-C., Batut J.,  
 RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yen K.-C., Batut J.,  
 RT "The composite genome of the legume symbiont Sinorhizobium meliloti."  
 RL Science 293:668-672(2001).  
 DR EMBL; AF155772; AAF27309.1; -  
 DR EMBL; AL591788; CAC46252.1; -  
 DR InterPro: IPR000733; flavo.monooxygenase.  
 DR Pfam: PF01360; Monooxygenase.  
 DR PRINTS; PR00420; RINGMONOXGNASE.  
 KW Complete proteome.  
 SQ SEQUENCE 425 AA; 45928 MW; E6EBCD75F0DE0FB7 CRC64;

Query Match 1.6%; Score 9; DB 16; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 ITGAGLAGS 94  
 |||||  
 Db 30 ITGAGLAGS 38

RESULT 8  
 ID 068141 PRELIMINARY; PRT; 445 AA.  
 AC 068141;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE GID PROTEIN.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 NC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;  
 CC Rhodospirillum.  
 CX NCBI\_TaxID=1061;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SB1003;  
 RX MEDLINE=97404404; PubMed=9256491;  
 RA Vitek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fomstein M.,  
 RT "Sequence of a 189-kb segment of the chromosome of Rhodospirillum  
 RT capsulatus SB1003."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).  
 DR EMBL; AF010496; AAC16231.1; -  
 DR InterPro: IPR002218; GIDA.  
 DR Pfam: PF003738; GIDA.1.  
 SQ SEQUENCE 445 AA; 47890 MW; 1C699BDC1EDD24C4 CRC64;

Query Match 1.6%; Score 9; DB 2; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 4.9;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 ITGAGLAGS 94

Db 6 ITGAGLAGS 14  
 |||||  
 RESULT 9  
 ID 082752 PRELIMINARY; PRT; 502 AA.  
 AC 082752;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE HYPOTHETICAL 55.2 KDA PROTEIN.  
 GN F7H9.220 OR AT4623030.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Bancroft I.,  
 RA Mewes H.W., Mayer K.F.X., Schueller C., Beyan M.,  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Mewes H.W.,  
 RA Lemcke K., Mayer K.F.X.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL031018; CA19819.1; -  
 DR EMBL; AL161558; CAB79258.1; -  
 DR InterPro: IPR002528; UPF0013.  
 DR Pfam: PF01554; UPF0013; 2.  
 KW Hypothetical protein.  
 SQ SEQUENCE 502 AA; 55232 MW; 3D5F8F6220B97BC4 CRC64;

Query Match 1.6%; Score 9; DB 10; Length 502;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 LLLCSLPIS 19  
 |||||  
 Db 122 LLLCSLPIS 130  
 RESULT 10  
 ID 049007 PRELIMINARY; PRT; 134 AA.  
 AC 049007;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE SIMILAR TO NADH OXIDASE.  
 OS Mycoplasma capricolum.  
 NC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 CC Entomoplasmataceae.  
 CX NCBI\_TaxID=2095;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 27343(KID);  
 RX MEDLINE=96059641; PubMed=7476192;  
 RA Bork P., Ouzounis C., Casari G., Schneider R., Sander C., Dolan M.,  
 RA Gilbert W., Gillevet P.M.,  
 RT "Exploring the Mycoplasma capricolum genome: a minimal cell reveals  
 RT its physiology."  
 RL Mol. Microbiol. 16:955-967(1995).  
 DR EMBL; Z33089; CAA83753.1; -  
 SQ SEQUENCE 134 AA; 14933 MW; 759690580E8F8DB0 CRC64;



Query Match 1.4%; Score 8; DB 2; Length 134;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAG 93  
|||||  
Db 7 IIGAGLAG 14

## RESULT 11

Q9PKF9 PRELIMINARY; PRT; 137 AA.  
AC Q9PKF9;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DE CYCLIC NUCLEOTIDE-BINDING PROTEIN, PUTATIVE.  
GN TC0506.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MOPN / NIGG;  
RX MEDLINE-20150255; PubMed-10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R.,  
RA Eissen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia  
RT pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL: AE002319; AAF39348.1; -;  
DR TIGR: TC0506; -;  
DR InterPro: IPR000595; CNMP\_binding.  
DR Pfam: PF00027; CNMP\_binding; 1.  
DR SMART: SMO0100; CNMP; 1.  
DR PROSITE: PS50042; CNMP\_BINDING\_3; 1.  
KW Complete proteome.  
SQ SEQUENCE 137 AA; 15410 MW; 607B533A70C9B7A7 CRC64;

Query Match 1.4%; Score 8; DB 16; Length 137;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 331 VALALLEL 338  
|||||  
Db 120 VALALLEL 127

## RESULT 12

Q970S5 PRELIMINARY; PRT; 173 AA.  
AC Q970S5;  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DE PUTATIVE NADH-UBIQUINONE OXIDOREDUCTASE SUBUNIT K.  
GN SP1527.  
OS Sulfolobus tokodaii.  
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
OX NCBI\_TaxID=111955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JCM 10545 / 7;  
RX PubMed-11572479;  
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
RA Sekine M., Baba S.-I., Ankal A., Kosugi H., Hosoyama A., Fukui S.,  
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,

RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
RA Oshima T., Kikuchi H.;  
RT "Complete genome sequence of an aerobic thermoacidophilic  
RT Crenarchaeon, Sulfolobus tokodaii strain7.";  
RL DNA Res. 8:123-140(2001).  
DR EMBL: AP000986; BAB6598.1; -;  
KW Ubiquinone; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 173 AA; 19125 MW; E7D964FB948939A1 CRC64;

Query Match 1.4%; Score 8; DB 17; Length 173;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 LPSEIGIP 313  
|||||  
Db 129 LPSEIGIP 136

## RESULT 13

Q97P95 PRELIMINARY; PRT; 197 AA.  
AC Q97P95;  
DT 01-OCT-2001 (TREMblrel. 18, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DE HYPOTHETICAL PROTEIN SP1746.  
GN SP1746.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-TIGR4;  
RX MEDLINE-21357209; PubMed-11463916;  
RA Tettein H., Nelson K.E., Paulsen I.T., Eissen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapfe E., Khouli H., Wolf A.M., Uitterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RL Science 293:498-506(2001).  
DR EMBL: AE007467; AAK75822.1; -;  
DR TIGR: SP1746; -;  
DR InterPro: IPR002819; HD.  
DR InterPro: IPR003607; HDC.  
DR Pfam: PF01966; HD; 1.  
DR SMART: SM00471; HDC; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 197 AA; 22548 MW; 7C6AD881F67805CB CRC64;

Query Match 1.4%; Score 8; DB 16; Length 197;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 LHDYAKKL 490  
|||||  
Db 57 LHDYAKKL 64

## RESULT 14

Q54770 PRELIMINARY; PRT; 206 AA.  
AC Q54770;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE HYPOTHETICAL 22.4 KDA PROTEIN (FRAGMENT).  
 OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PCC 7942;  
 RA Phung L.T., Haselkorn R.;  
 RT "Genes encoding the alpha subunit of carboxyltransferase of the  
 acetyl-coA carboxylase complex and GTP cyclohydrolase I from  
 cyanobacterium Synechococcus sp. PCC 7942."  
 RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: U59236; AAB82044.1; -.  
 DR InterPro: IPR003719; PhzC\_PhZF.  
 DR Pfam: PF02567; PhzC-Phzf.1.  
 KW Hypothetical protein.  
 FT NON\_TER 206  
 SQ SEQUENCE 206 AA; 22386 MW; 9F2EE53736A97C00 CRC64;

Query Match 1.4%; Score 8; DB 2; Length 206;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 268 TLAAQAAL 275  
 |||||  
 DB 75 TLAAQAAL 82

RESULT 15  
 Q9BJ45  
 ID Q9BJ45 PRELIMINARY; PRT; 323 AA.  
 AC Q9BJ45;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 36.6 KDA PROTEIN.  
 OS Leishmania major.  
 OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21256109; PubMed=11356515;  
 RA Pedrosa A.L., Ruiz J.C., Tosi L.R.O., Cruz A.K.;  
 RT "Characterization of three chromosomal ends of Leishmania major  
 RT reveals transcriptional activity across arrays of reiterated and  
 RT unique sequences."  
 RL Mol. Biochem. Parasitol. 114:71-80(2001).  
 DR EMBL: AF339905; AKK28283.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 323 AA; 36566 MW; 8C92C045309068DF CRC64;

Query Match 1.4%; Score 8; DB 5; Length 323;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ADPLQQL 115  
 |||||  
 DB 102 ADPLQQL 109

Search completed: August 20, 2002, 04:30:01  
 Job time: 280 sec



**THIS PAGE BLANK (USPTO)**

PSSPQLATEKRVSLPESLFEKHEVRMTIAKLIENGKAKRTIKRSLSTIKSFAHYC  
 VIOKILLENPAETIHGPRLPELSPMTYAOVETAMTPDISKPHGJLDRCLMELFYS  
 SGLRIEIVANMODFDLSLFLIRIROGKKEKRIIPTYSNAIOMIOIYLNHMDKRLIE  
 KDPAIFLNRGRIRISIDRSFOEYLRRSGSLGSHIPIHRTIAFHMLESDMLK  
 TIOALLHSHSELTYYTQVSVRKAKROTDBAHPHA\*  
 BASE COUNT 4710 a 3330 c 3665 g 4743 t  
 ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 16448;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtcgctgatttggaaaaagtcagtgatggttataatgctcaagctccatcc 60  
 Db 13290 GTGGCTGATTTTGAAGAAAGCCATGATGTTTATATATTGAAGTCCCTATCC 13231  
 QY 61 aaacattgaaactgtctagagaggttgaaatgatactatgaggaactatccatac 120  
 Db 13230 AAACATTGAATATCTGTAGAGAGATTGAACATGATCTATGAGACTATCTATAC 13171  
 QY 121 tctcttggaactttattgtgtatgcttccatcttctctgtgttaattccctgagtc 180  
 Db 13170 TCTCTTGGACTTTATGTGTAGTCTCCATTTCTGTGTGCTAAATTCCTGAGTC 13111  
 QY 181 tgaagtcataagatcccttatataagtcagcaatctacacagcagccttagaacaata 240  
 Db 13110 TGAAGTCAATAAGATCCTTATATATAGTACGCAATCTACACAGCAGCTTAGCAACATA 13051  
 QY 241 tctggaagctcagatgctcagtgatcatgactctctgtttaaagaaaaatcgagaa 300  
 Db 13050 TCTGGAAGCTTATAGTCTAGAGTATCATGACTCTTGTTTAAAGAAATCGGAGA 12991  
 QY 301 agactatctcaagcaaaagatccactccctcagatccgaaacttagaaaaagacatcat 360  
 Db 12990 AGACTATCTCAAGCAAAAGATCCACTCCCAATCCGAAACTAGAAAAAGACCATCAT 12931  
 QY 361 tggagagagcctgagcgagatcttcaagaagccttgagcgtgctcccaagcatatgaaac 420  
 Db 12930 TGGAGAGAGCCTGGCGGATCTTCAGAAAGCCTTGAGCGTCTCCCAAGCATATGGAAC 12871  
 QY 421 tgcagagccctgacagcagctactggtttatcagcagctcagaaacatcttgggaaac 480  
 Db 12870 TGCAGAGCCCTGACAGCAGCTACTGTTTATCGGCAAGTCTGAGCATCTTGGGAAAC 12811  
 QY 481 ttctgagcactactgtttaaagctttagatctcccttccctgtatccgttagaaac 540  
 Db 12810 TTCTGAGCACTACTGTTTAAAGCTTTAGCATCTCCCTATCCGTATCCGCTTAGAAGC 12751  
 QY 541 cgcctatagactgtcatttgaagaaacataagtcattgatactatcatcttcat 600  
 Db 12750 CGCCTATAGACTGTCTATTTGAAGAACTAAAGTCAATTGATCATCTATCTTTCAAT 12691  
 QY 601 tcaatagctcccgaaagaaatccaatgctatctgcggaataatctcagcttggagac 660  
 Db 12690 TCATTAAGCTTCCGGAAGAAATCCAAATGCTATCTGCGCAATATCTTCAAGCTTGAAGAC 12631  
 QY 661 tgaagaatcagatgcttatatcgggactcttagctgcgaagaaagcggaattggag 720  
 Db 12630 TGAAGAATCTGATGCTTATATTCGGGATCTTAACTGCGCAAAAGGCGGATTCGGAG 12571  
 QY 721 tgcacagacttgcagatcaggaataacacaaagaaagcttcttcgcagaaacttagaa 780  
 Db 12570 TGCACAGACTTTCAGATTCGGAGAAATACCAACAAAGCTTCTTCCGACACTTATGGA 12511  
 QY 781 ttgtgataagatgctgtctctcctaagatcaagaagctatctttagctttaggaagct 840  
 Db 12510 TTTGTAAACAGATGCGCTCTCTCAAGATCAAGAAAGCTATCTTATGCTTAGGGAAGCT 12451  
 QY 841 taagagatgctcagagctactacataataaaagaaagcaattgcagaagccttagatgt 900  
 Db 12450 TAAAGATGCTCAGAGCTACTACATATTAATAAAAGCAATTCAGAAAGCCTGATGTGATGT 12391

QY 901 caattagcagcgtcagacttaattgtcttgggaaagaaagcgtcttccgt 960  
 Db 12390 CACTTAGCAGCAGCTCAAGCTTAATGCTTTGGGAAAAAGAGAGCGCTTCCGT 12331  
 QY 961 gataaagaagcagcacttggagcgccctcgagccctgtatgctttaggaatcacc 1020  
 Db 12330 GATAAAGAGCAGCACTTGAGAGCGCCCTCGAGCCCTGTATCTTACGCTATAC 12271  
 QY 1021 ctctgagaatggatctcgatcttgcctcgatctcgaattctcctaaagaaagcagcgaagc 1080  
 Db 12270 CTCTGAGTATGAGATTCGATTCGCCCTCGCATATCTTAATAAGTAAAGCAGAGC 12211  
 QY 1081 caagttgaatgactttagctctctcttagagttgaagtggtgtacaccccttaacttga 1140  
 Db 12210 CAAGTTGAATGAGCTTAGCTCTCTTAGAGTATGAGGTGAGACCCCTTAAGTACTGA 12151  
 QY 1141 ataatcagaaagcgttgcacacacatataatgaagactcagccttagatcttc 1200  
 Db 12150 ATACATTAACGGAAGCTGTGTCACACCATATATATAGACTCTGAGCTTGAATCTC 12091  
 QY 1201 taagggggtactcttaaaaaattggaagcggtgagacataatgctccctaaagatcccca 1260  
 Db 12090 TAAAGGGGTACTCTTAATAAATTGGAAGCGGTTGAACATCATAGTCCCTCAAGATCCCA 12031  
 QY 1261 ggaaggggaaaggttgcctccacacacagaggtcttgaagagcagatccttagct 1320  
 Db 12030 GGAGAGGGAAAGCTTGCTCTCCACAAACCGAGGCTTGAAGAGCAGATCTTAGCTTCT 11971  
 QY 1321 ctccgctacctaagaagcttaccctccctgtatataagcttggcgagtcagaa 1380  
 Db 11970 CTTCGCGCTACTTAAGAAAGCTTACCTCCCTGTTTATATAGCTTTGGGAGTCAAA 11911  
 QY 1381 aactagcttgcagactacagatctctttaaagtaaacactacatacgaagagcctt 1440  
 Db 11910 AACTAGCTTGCACACTACTGCGATTTCTTTTAAAGTACACACTTACATCAGAGAGCCTT 11851  
 QY 1441 agatctacttctcagaagcttgcagagcttctcgaagaaactatcatccgcgactatgaca 1500  
 Db 11850 AGATCTACTTTCGAAGCTGCGAGCTCTCTGGAAGACCTATCATCCGCGCTATGACGA 11791  
 QY 1501 tcttgcattatataatctcaacaaagatccctgaaaaaaagcttctccatgatatgc 1560  
 Db 11790 TCTTCTATATTTAATCTCACCAAGATCTGAAAAAAACCTTCTCCATGATTATGC 11731  
 QY 1561 aaaaaagcctaattcagaaagccttcttatttggaaacggaaacaaagacccatcc 1620  
 Db 11730 AAAAAAGCTAATTCAGAAACCTTGTATTGTGACACGGAACCAAGACCCCATCC 11671  
 QY 1621 cagcatgcccatactcagttatcaggtcaccacaaagaaagcgttacgaagctcatgtgga 1680  
 Db 11670 CAGCATGCCCTATCTACGTTATCAGGTACCCACAGAAAGCCGTAGAGCTCATGTGGA 11611  
 QY 1681 tattctagaagacactcagcactcgaggtcttcggaagatacgtttatgatacaact 1740  
 Db 11610 TATTCTAGAGACACTAGCCACTCGAAGCTTCGAAAGTATCCGTTATATATCAACT 11551  
 QY 1741 gatgacggaagagatgacaaaaaattccagcttcccttgaagcttaccataaaatgt 1800  
 Db 11550 GATGACGGAAGAGATGCAAAAAATTTCCCACTCTTGAAGCTTACATATAAATGT 11491  
 QY 1801 ggaagtaaccccaactcagctctatagaaagcttgccttattcttagcttccctgttc 1860  
 Db 11490 GGAGTAACCCCAACTACGCTTATGAACGTTGCTTATTTAGCTTCTTGTTC 11431  
 QY 1861 ttatggttccctcagctgagctttagactcacaagaagctgtgaa 1907  
 Db 11430 TTATGGTTCCTCAGCTGATGCTTGAACATCAAGAGGCTGTGAAA 11384

RESULT 4  
 AX349527 1707 bp DNA linear PAT 06-FEB-2002  
 LOCUS AX349527  
 DEFINITION Sequence 50 from Patent WO0202606.

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:10:16 ; Search time 66.94 Seconds  
(without alignments)  
1467.898 Million cell updates/sec

Title: US-09-662-812-2  
Perfect score: 2833  
Sequence: 1 MGLEHLTLFGLLCSLPISL.....TEGDKNPVLGLIKIVE 568

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SEPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2833	100.0	568	16	Q929G0 chlamydia p
2	1456	51.4	566	16	Q9PK44 chlamydia m
3	1421	50.2	566	16	Q84354 chlamydia t
4	136.5	4.8	514	5	Q9UB78 drosophila
5	134.5	4.7	514	5	Q76522 drosophila
6	134	4.7	860	3	P87295 schizosacch
7	133	4.7	321	16	Q91112 pseudomonas
8	133	4.7	1138	16	Q84133 chlamydia t
9	132.5	4.7	514	5	Q9V455 drosophila
10	132.5	4.7	948	4	Q9UK7 homo sapien
11	132.5	4.7	1200	17	Q9HKM6 thermoplasma
12	132.5	4.7	2554	2	Q30981 bacillus su
13	131	4.6	398	2	Q44295 anabaena sp
14	130.5	4.6	979	11	Q9EP71 mus musculu
15	129.5	4.6	1316	10	Q9M6P9 arabidopsis
16	129.5	4.6	1973	3	Q9Y7K2 schizosacch

17	129.5	4.6	7201	5	Q9VMT8 drosophila
18	127.5	4.5	1978	4	Q15154 homo sapien
19	126.5	4.5	1292	10	Q9SL02 arabidopsis
20	124.5	4.4	526	16	Q9PK10 chlamydia m
21	124	4.4	712	11	Q9CRX4 mus musculu
22	124	4.4	1211	11	Q9R0L3 rattus norv
23	124	4.4	1235	11	Q9R0L5 rattus norv
24	124	4.4	1273	11	Q9R0L4 rattus norv
25	124	4.4	1275	3	Q9PE64 neurospora
26	123.5	4.4	890	10	Q9M649 arabidopsis
27	123.5	4.4	894	10	Q9S0S3 arabidopsis
28	123.5	4.4	894	10	Q9M650 arabidopsis
29	123.5	4.4	1220	3	Q9P3A8 schizosacch
30	123.5	4.4	1591	4	Q92601 homo sapien
31	123.5	4.4	1723	2	Q9JMX8 heliobacte
32	123.5	4.4	1979	4	Q9S549 homo sapien
33	122	4.3	719	10	Q81742 arabidopsis
34	122	4.3	1318	4	Q43254 homo sapien
35	121.5	4.3	1841	4	Q15031 homo sapien
36	121	4.3	596	16	Q9RVT5 deinococcus
37	121	4.3	846	10	Q9EP99 oryza sativ
38	120.5	4.3	430	17	Q9UT9 pyrococcus
39	120.5	4.3	937	11	Q9DBG3 mus musculu
40	120.5	4.3	951	4	Q96019 homo sapien
41	120	4.2	922	16	Q9HMA7 pseudomonas
42	120	4.2	1388	6	Q9GL21 canis fami
43	119.5	4.2	669	4	Q96M49 homo sapien
44	119.5	4.2	1466	10	Q9ZCO3 arabidopsis
45	119.5	4.2	1588	11	Q9ESK9 mus musculu

## ALIGNMENTS

RESULT	ID	Q929G0	PRELIMINARY;	PRT;	568 AA.
AC	Q929G0	Q929G0			
DT	01-MAY-1999	(TEMBREL. 10, Created)			
DT	01-MAY-1999	(TEMBREL. 10, Last sequence update)			
DT	01-DEC-2001	(TEMBREL. 19, Last annotation update)			
DE	POTATIVE OMP PROTEIN.				
GN	CN00021 OR CPJ0021 OR CP0755.				
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
OX	NCBI_TaxID=83558;				
RN	[1]	SEQUENCE FROM N.A.			
RP	STRAIN=CWL029;				
RC	MEDLINE=9206606; PubMed=10192388;				
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,				
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.,				
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."				
RL	Nat. Genet. 21:385-389(1999).				
RN	[2]	SEQUENCE FROM N.A.			
RP	STRAIN=AR39;				
RC	MEDLINE=20150255; PubMed=10684935;				
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,				
RA	White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Bass S.,				
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,				
RA	Gavin M., Nelson J., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,				
RA	Eisen J., Fraser C.M.;				
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia				
RT	pneumoniae AR39."				
RL	Nucleic Acids Res. 28:1397-1406(2000).				
RN	[3]	SEQUENCE FROM N.A.			
RP	STRAIN=J138;				
RC	MEDLINE=20330349; PubMed=10871362;				
RA	Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,				
RA	Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;				
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138				

RT from Japan and CML029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AE001587; AAD18174.1;  
 DR EMBL; AE002234; AAF38558.1;  
 DR EMBL; AP002545; BAA98233.1;  
 DR TIGR; CP0755;  
 DR InterPro; IPR004155; HEAT\_PBS.  
 DR Pfam; PF03130; HEAT\_PBS.  
 KW Complete proteome.  
 SQ SEQUENCE 568 AA; 63553 MW; 3358FCB5BC482E80 CRC64;

Query Match 100.0%; Score 2833; DB 16; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-179;  
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLFHLTFGLLCSPISLVAKPPESYGKILYISTOSTOQALATYIEALDAYGDHDF 60  
 DB 1 MGLFHLTFGLLCSPISLVAKPPESYGKILYISTOSTOQALATYIEALDAYGDHDF 60  
 QY 61 VLKRIGEDYKQSIHSSDPQTRKSTIIIGAGLAGSSEALDVLISQAMETADPLQQLLVLSAV 120  
 DB 61 VLKRIGEDYKQSIHSSDPQTRKSTIIIGAGLAGSSEALDVLISQAMETADPLQQLLVLSAV 120  
 QY 121 SGHLGKTSDDLFFKALASPYVIRLEAAYRLANKNTVYIDHLSFIKHPBEIOCLSA 180  
 DB 121 SGHLGKTSDDLFFKALASPYVIRLEAAYRLANKNTVYIDHLSFIKHPBEIOCLSA 180  
 QY 181 IFLRLETESDAYIRDLAALAKKSAIRSATALQIGEQKRLPTLRNLITSASPODOBAI 240  
 DB 181 IFLRLETESDAYIRDLAALAKKSAIRSATALQIGEQKRLPTLRNLITSASPODOBAI 240  
 QY 241 LYLAKGLKDGOSYNIKKLOKPDVDTLAAQAALIALGKEEDALPVIKQALEERPRAL 300  
 DB 241 LYLAKGLKDGOSYNIKKLOKPDVDTLAAQAALIALGKEEDALPVIKQALEERPRAL 300  
 QY 301 YALHRLPSEIGIPALPFLTKNSEAKLNVALLELGCDTPKLEVITTEKVOAPHNE 360  
 DB 301 YALHRLPSEIGIPALPFLTKNSEAKLNVALLELGCDTPKLEVITTEKVOAPHNE 360  
 QY 361 TLALSFSKGRLLQNMKNRNIIIVPODERERLLSTRGLEQILTFELRLEKREAVLPCTY 420  
 DB 361 TLALSFSKGRLLQNMKNRNIIIVPODERERLLSTRGLEQILTFELRLEKREAVLPCTY 420  
 QY 421 KLASQKQTALATTAISFISHTSHOEALDLFQAAKLPGEPITIRAYADLAIVNLTDPPEKK 480  
 DB 421 KLASQKQTALATTAISFISHTSHOEALDLFQAAKLPGEPITIRAYADLAIVNLTDPPEKK 480  
 QY 481 RSLHDYAKKLIQETLLFVDTENORPHPSMRYLRYQVPESTRKMLDLLETFLATSKSSED 540  
 DB 481 RSLHDYAKKLIQETLLFVDTENORPHPSMRYLRYQVPESTRKMLDLLETFLATSKSSED 540  
 QY 541 IRLLIQMTGEDAKNFVLAGLIIKIVE 568  
 DB 541 IRLLIQMTGEDAKNFVLAGLIIKIVE 568

RESULT 2  
 Q9PK44 PRELIMINARY; PRT; 566 AA.  
 AC Q9PK44;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOHETICAL PROTEIN TC0629.  
 GN TC0629.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_Taxid=83360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MOPN / NING;  
 RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBooy R., Kolonay J., McClarty G., Salberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
 RT pneumoniae AR39."  
 RT Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE002331; AAF39458.1;  
 DR TIGR; TC0629;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 566 AA; 63143 MW; F9C7B8E7B4E1C6C2 CRC64;

Query Match 51.4%; Score 1456; DB 16; Length 566;  
 Best Local Similarity 53.5%; Pred. No. 3e-88;  
 Matches 304; Conservative 92; Mismatches 170; Indels 2; Gaps 1;

QY 1 MGLFHLTFGLLCSPISLVAKPPESYGKILYISTOSTOQALATYIEALDAYGDHDF 60  
 DB 1 MGLFHLTFGLLCSPISLVAKPPESYGKILYISTOSTOQALATYIEALDAYGDHDF 60  
 QY 61 VLKRIGEDYKQSIHSSDPQTRKSTIIIGAGLAGSSEALDVLISQAMETADPLQQLLVLSAV 120  
 DB 61 VLKRIGEDYKQSIHSSDPQTRKSTIIIGAGLAGSSEALDVLISQAMETADPLQQLLVLSAV 120  
 QY 121 SGHLGKTSDDLFFKALASPYVIRLEAAYRLANKNTVYIDHLSFIKHPBEIOCLSA 180  
 DB 121 SGHLGKTSDDLFFKALASPYVIRLEAAYRLANKNTVYIDHLSFIKHPBEIOCLSA 180  
 QY 181 IFLRLETESDAYIRDLAALAKKSAIRSATALQIGEQKRLPTLRNLITSASPODOBAI 240  
 DB 181 IFLRLETESDAYIRDLAALAKKSAIRSATALQIGEQKRLPTLRNLITSASPODOBAI 240  
 QY 241 LYLAKGLKDGOSYNIKKLOKPDVDTLAAQAALIALGKEEDALPVIKQALEERPRAL 300  
 DB 241 LYLAKGLKDGOSYNIKKLOKPDVDTLAAQAALIALGKEEDALPVIKQALEERPRAL 300  
 QY 301 YALHRLPSEIGIPALPFLTKNSEAKLNVALLELGCDTPKLEVITTEKVOAPHNE 360  
 DB 301 YALHRLPSEIGIPALPFLTKNSEAKLNVALLELGCDTPKLEVITTEKVOAPHNE 360  
 QY 361 TLALSFSKGRLLQNMKNRNIIIVPODERERLLSTRGLEQILTFELRLEKREAVLPCTY 420  
 DB 361 TLALSFSKGRLLQNMKNRNIIIVPODERERLLSTRGLEQILTFELRLEKREAVLPCTY 420  
 QY 421 KLASQKQTALATTAISFISHTSHOEALDLFQAAKLPGEPITIRAYADLAIVNLTDPPEKK 480  
 DB 421 KLASQKQTALATTAISFISHTSHOEALDLFQAAKLPGEPITIRAYADLAIVNLTDPPEKK 480  
 QY 481 RSLHDYAKKLIQETLLFVDTENORPHPSMRYLRYQVPESTRKMLDLLETFLATSKSSED 540  
 DB 481 RSLHDYAKKLIQETLLFVDTENORPHPSMRYLRYQVPESTRKMLDLLETFLATSKSSED 540  
 QY 541 IRLLIQMTGEDAKNFVLAGLIIKIVE 568  
 DB 541 IRLLIQMTGEDAKNFVLAGLIIKIVE 568

RESULT 3  
 Q9PK44 PRELIMINARY; PRT; 566 AA.  
 AC Q9PK44;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE HYPOHETICAL 63.5 KDA PROTEIN.  
 GN C7350.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_Taxid=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.



RC STRAIN-D/UM-3/CX;  
MEDLINE-9900809; PubMed-9784136;  
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.,  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis";  
Science 282:754-759(1998).  
DR EMBL: AE001308; AAC67945.1;  
DR InterPro: IPR004155; HEAT\_PBS.  
DR Pfam: PF03130; HEAT\_PBS; 5.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 566 AA; 63507 MW; 92E5DAF190B23D56 CRC64;

Query Match 50.2%; Score 1421; DB 16; Length 566;  
Best Local Similarity 53.0%; Pred. No. 6.1e-86;  
Matches 301; Conservative 88; Mismatches 177; Indels 2; Gaps 1;

OY 1 MGSHLTLFGLICSLPISLAKFPESVGHKILYISTOSTQALATVYLAIDAYGDHDF 60  
D 1 MGSLRIAFISFL--SFTLSASCDPSSVSQRILFSCSKSVPALEATLEASATYQCHDES 58  
OY 61 VLRKIEDVYKOSIHSSDPQFRSTTIAGLAGSSSEALDVLQAMETADPLQQLVLSAV 120  
D 59 VLRVIESYLDQSFLESDYIRKSAIIGAGLSSSEALIELSEAIETQDLYEQLILNMA 118  
OY 121 SGHLGTSDDLFLKALASPVYIRLEAAYRLAKNTKYVDHLSFHKRPEEIOCSAA 180  
D 119 TQSLSTSDLLFKGLTASHPIRLAAYRLAKMKNKSVSDLYSFYKPEEYQNLAA 178  
OY 181 IFRLTEESDAYIRDLAKKSAIRSATLQGEYQOKRFLPTLRNLTSASPODEAI 240  
D 179 IFQLETEEDAIHHLSSPNLTRYAVYLIGETKQKFLPTLSLSASPLDEGA 238  
OY 241 LVALGKIDGQSYNNIKQLOKPDVDTLAAQALIALGKEEDALPVYKQALEERPRL 300  
D 239 LVALGKLEDSGYRIKALSSRSNPVYLAQAQTLFLKEKEEALPILTNLCOQKLRL 298  
OY 301 YALRHLPSELGIRIAPILPKTKNSEAKLVALALEGCDTRKLEYTERLYQPYNE 360  
D 299 YARFELSQEGEELLPIEFNATQEBIRNTALVALHQGTDPOVLYLLEIESKVLH 358  
OY 361 TLALSESKGTLQNMKNVNIIVPODERERLSTTGLEOILTFRLPKREAYLPCY 420  
D 359 IFLPTSTGTAIOFWKECTTFPLMSQEDKMTAMRYVADTILSLKLPNDAYLYLE 418  
OY 421 KLLASQKQTLATTAISFLSHTSHQEAALLFOAKLPGEPIIRAYADLAITYNLKPEEK 480  
D 419 RLASQKTLIAKAIAFLSVTAHQALSLYSKALPFGDPIIRAYALALYTMKDEEK 478  
OY 481 RSHDAKAKLIOETLFLVTEENORPHSPMYLKYOTPEERTKLMIDLTLATSKSED 540  
D 479 AVLYRAAEIETDITLFTAEALNPSSSSYLRYOVSPETRTQMLALILETVSSKTED 538  
OY 541 IRLLIQMTGDAKNFVLALIKIYE 568  
D 539 IRYFLSMKTKHYKNPIISGLMRYVE 566

RESULT 4  
ID 09UB78 PRELIMINARY; PRT; 514 AA.  
AC 09UB78:  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE IMPORTIN ALPHA-3 (IMPORTIN ALPHA 3).  
GN KARYOPHERIN-ALPHA-3 OR KARYOPHERIN-ALPHA3 OR CG9423.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mathe E., Bates H., Huikeshoven H., Glover D., Cottrell S.,  
RT "Importin- $\alpha$ 3 is required at multiple stages of Drosophila development  
RT and has a stage specific role in the completion of oogenesis";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Mathe E., Cottrell S.,  
RT "Importin- $\alpha$ 3 cDNA (gm06753), Drosophila melanogaster";  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Mathe E., Cottrell S.,  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ37997; CAB40789.1;  
DR EMBL: AF230873; AAF37856.1;  
DR EMBL: AF230872; AAF37855.1;  
DR EMBL: AF230871; AAK14941.1;  
DR HSSP: 002821; 1BK5.  
DR FlyBase: FBgn0027338; karyopherin- $\alpha$ 3.  
DR InterPro: IPR000225; Armadillo.  
DR InterPro: IPR002652; IBB.  
DR Pfam: PF00514; Armadillo\_seg; 8.  
DR Pfam: PF01749; IBB; 1.  
DR SMART: SM00185; ARM; 6.  
DR PROSITE: PSS0176; ARM\_REPEAT; 2.  
SQ SEQUENCE 514 AA; 57036 MW; 5F007D9C73B85B8F CRC64;

Query Match 4.8%; Score 136.5; DB 5; Length 514;  
Best Local Similarity 21.1%; Pred. No. 0.52; Indels 155; Gaps 23;  
Matches 108; Conservative 70; Mismatches 179;

OY 95 SEALDVLQAMETAD---PLQQLVLSAVSGHLGKTSDDLFLKALASPVYIRLEAAYRL 151  
D 62 SSSIDLKLLAKKAADATKPEQQLAA-----YQAARKL 93  
OY 152 ANL-KNTKYVDHLSFHKRPEEIOCSAALFLRLEESDAYIRDLAKKSAIRSATA 210  
D 94 LSLDKNPINDLIQSDI--PLIIVELCKOHNTMLQFEAAWAL-----TNIASTGS 142  
OY 211 LQIGEYQOKRFLPTLRNLTSASPODEALITYALGK-----LKDGQSYNNIKQLO-- 261  
D 143 AQINEVVAAGAVPLFDLQNSPAPNCEQAWALGNIIIGDPLRDVYIKRGVQPLLSF 202  
OY 262 -KPDVDTLA--AAQALIALGKEEDALPVYKQALEERPRLAYLRH----- 305  
D 203 IKRPDIPITFLRNWVWVYVNLCKRNDPAP--PTATIHETLRALNLTHTDNLIVDTYWA 260  
OY 306 -----LPSGIG-IPLALPIFLTKNSEAKLVN---ALALLEIGCDT----- 342  
D 261 ISYLTGDNQDQIOWIVIESGVVPLIPL--GNSEVVRQVTAALAVGNITVYGSDEQOVV 317  
OY 343 -----PKLEYITERLYQPHYNETLASEKSGTLQNMKNVNIIVP----- 383  
D 318 LNYDALSYFPELLSHPEKRIKKE-----AWFLSNITAGQSOVOAVINGLPKIIEEN 371  
OY 384 -----ODPOERERILS--TTRGLEOILTFRLPKREAYLPCYIKLASQKQTLATTAI 435  
D 372 LSKGEFQTKQEAAMAINLNTISGREGVFTLL--KEGVIPPCDLSQDQVINYVL 427  
OY 436 SFLSH-----TSHQAL-DLLFOAKLPGEPIIRAYADLAITYNLKPEKKRSJHDYAKK 489  
D 428 DGLNLMKLVADSHVEAVANCIIECEGLAKIERLOSHEWVEYKL-----AYE 474  
OY 490 LIOETLFLVDT-ENQRPHPSPMYLRQVTPES 520  
D 475 IIDQ--YFTDGEQOTNMAPSSDGAQYINFDPA 504

RESULT 5

ID	076522	PRELIMINARY;	PRN:	514	AA.
AC	076522;				
DT	01-NOV-1998	(TREMBLrel, 08, Created)			
DT	01-NOV-1998	(TREMBLrel, 08, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel, 17, Last annotation update)			
DE	KARYOPHERIN ALPHA 3 (FRAGMENT).				
GN	KARYOPHERIN-ALPHA-3 OR CG9423.				
OS	Drosophila melanogaster (fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydridae; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Mason D.A., Goldfarb D.S.;				
RT	"Drosophila melanogaster karyopherin alpha 3.";				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF0714958; AAC26056.1; -.				
DR	HSSP; Q02821; 1BK5.				
DR	FLYBase; FBgn0027335; karyopherin-alpha-3.				
DR	InterPro; IPR000225; Armadillo.				
DR	InterPro; IPR002652; IBB.				
DR	Pfam; PF00514; Armadillo_seg; 8.				
DR	Pfam; PF01749; IBB; 1.				
DR	SMART; SM00185; ARM; 6.				
DR	PROSITE; PSS0176; ARM_REPEAT; 2.				
FT	NON_TER	514	514		
SQ	SEQUENCE	514	AA;	56975	MM; 78500AC4EA906D3A CRC64;

Query Match	4.7%	Score 134.5	DB 5	Length 514
Best Local Similarity	20.9%	Pred. No. 0.71		
Matches 107	Conservative 69	Mismatches 182	Indels 153	Gaps 22

[illegible]

## RESULT 6

ID	P87295	PRELIMINARY:	PRT:	860 AA.
AC	P87295.			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	HYPOTHETICAL 99.0 KDA PROTEIN C16A10.03C IN CHROMOSOME I. SPAC16A10.03C.			
GN	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OX	Schizosaccharomycetes.			
RN	NCBI_TaxID=4896;			
RP	[1]			
RC	SEQUENCE FROM N.A.			
SC	STRAIN=972;			
RA	Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;			
RL	Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- SIMILARITY: SOME. TO YEAST ENDI.			
DR	EMBL; Z97185; CAB09596.3; -			
DR	InterPro; IPR001841; Znf_Tfing.			
DR	SMART; SM00184; RING; 1.			
KW	Hypothetical protein; Transmembrane.			
FT	TRANSMEM 145 165			POTENTIAL.
FT	TRANSMEM 261 281			POTENTIAL.
FT	TRANSMEM 431 451			POTENTIAL.
FT	TRANSMEM 564 584			POTENTIAL.
SO	SEQUENCE 860 AA; 99031 MW; 52ABDC9ACCE72CCD CRC64;			

Query Match	4.78	Score 134	DB 3	Length 860
Best Local Similarity	18.48	Pred. NO. 1.6		
Matches 118	Conservative	85	Mismatches 234	Indels 206
				Gaps 19

[illegible]

0

Db 668 -DSLIOIISDPEVTKL-----SETYSESDALHYLKFFVERRSITNKY 1710  
QY 526 LDLETLATSKSSE---DIRLLIQIMTEGDAKNEPVLGLLTK 565  
Db 711 EDLYKILBACFMQFRIPQIOWHLNLLVKGDTLNFCLPLLLK 753

RESULT 7  
091112 PRELIMINARY; PRT; 321 AA.  
AC 091112;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA2293.  
GN PA2293.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garner R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yan Y.,  
RA Brody L.L., Collier S.N., Folger K.R., Kas A., Lardy K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reiser J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004653; AAC05681.1; -  
DR InterPro; IPR004155; HEAT\_PBS.  
DR InterPro; IPR000357; HEAT\_repeat.  
DR Pfam; PF03130; HEAT\_PBS; 2.  
DR PROSITE; PS50077; HEAT\_REPEAT; 1.  
KW Hypothetical protein: Complete proteome.  
SQ SEQUENCE 321 AA; 34028 MW; 20A07E919333DB0B CRC64;

Query Match 4.7%; Score 133; DB 16; Length 321;  
Best Local Similarity 25.5%; Pred. No. 0.45;  
Matches 76; Conservative 31; Mismatches 137; Indels 54; Gaps 10;

QY 70 LKOSHSDBQTRKSTTIGAGLAGSSBALDVLVSQAME-TADPLQQLVLVSAGHGLKTS 128  
Db 14 LSPRLDADPGVRRLLALIELADELPALPLVLAALRGDDP-----GVGGEARLL 65  
QY 129 D-----DLFLKALASPVVIRLEAAVRLANLKNKTV-----IDHLSPFHKLPPE 173  
Db 66 EAWEDAVNALCALADPVAADAAAGSLGELKEPACGRLLPWIGHADAV-----R 120  
QY 174 IQCLSAIFLRETEESDAYIRDLAAKSAISATATLQIGEYOQKRFLLRLMLTS-A 232  
Db 121 ASVLRAIRLERLT---EEGAVVALAALGDPOAAVREAAVAVGMLRHQALAEKLASADV 178  
QY 233 SPDDQEAFLYALGKLKDQSYINIKQLOKPDVDVTLAAQALIALGKEDALPVYIK-- 290  
Db 179 DPEVRRRAATGALISREATVLPALCALADAOQVREAAATTLGKLGREBAGEPLTKALA 238  
QY 291 -----QALEERPRALYALRHLPSE-----IGIPALPFIPLKTNSEKLVALLLEL 338  
Db 239 DDYWOVRLRARALGRLRHRRPAREALEALIGHPI-----GNLRKEAVALALGEL 286

RESULT 8  
084133 PRELIMINARY; PRT; 1138 AA.  
AC 084133;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE (POSSIBLE TRANSMEMBRANE PROTEIN).  
GN CT131.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/CX;  
RX MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
RT Chlamydia trachomatis.";  
RL Science 282:754-759(1998).  
DR EMBL; AE001286; AAC67722.1; -  
KW Transmembrane; Complete proteome.  
SQ SEQUENCE 1138 AA; 126751 MW; 583015347FE926A5 CRC64;

Query Match 4.7%; Score 133; DB 16; Length 1138;  
Best Local Similarity 19.9%; Pred. No. 2.8;  
Matches 158; Conservative 119; Mismatches 240; Indels 278; Gaps 41;

QY 8 LFG-LILCSPLISLVA---KEPESVGHKILYISTQSQOALATYLEALDA--YG----- 55  
Db 9 IFQYLLVGLVPLALLLPKFFSSSESGKYLFVLVNLKGTGLQPEIDELHLSWFSQYAKK 68  
QY 56 -----DHDFFVLKRGEDYLKQSIHSSDPQ-----TRKSTTIGAGLAGSS 95  
Db 69 IRIKIGDSDEIFAEEKI-----LYKGSPLRLLIRFPKALTLQWLSQIDBSLSMNS 121  
QY 96 EAL-----DVLQAMETADPLQOL--LVLSAVSG-----HLGKTSDDLKRALA--- 137  
Db 122 PSYHLDPGVILSKIERSDITSELGSIYTKMTNGSTLSVSGFYKKAEQDLRLALRKEN 181  
QY 138 -----SPYPIYRIE-----AAVRLNKNKTKYIDHLSPFHKLPPE 172  
Db 182 DVGESAVNEGALSPNVLNVELSVPASLFRKPLIASLDRILSTENLINTAKAHQEKD 241  
QY 173 EIQLSAIFLRETEESDA---YIRDLAANKSAISATATLQIGEYOQKRFLLRLN- 227  
Db 242 -----STLLITLTAEGNQSIAKRGYVDRAFLLITOGGASSVILQ-----PTTTSR 286  
QY 228 LILSASPOD-----QEAILY-----ALGKLKQSQSYINIKQLOK-----PDVDVTLAA 271  
Db 287 ILSELSPLDPPIRSQEAYLFISEAKLPLSISKWSADFSLOANLPQISVDPDPNLSIRT 346  
QY 272 AQLIALGKEDALPVYIKQ-----ALEERPR-----L 300  
Db 347 ENRKIST-RKSDHLTVIRSSSSAALGASPSYIHSITSMNKHVAEFHQLSHLPHRYL 405  
QY 301 YALNHPSEIGIPALPFL---KTKNSEAKLNVAL--ALLEAGDTP----- 343  
Db 406 RALLPKPIELNTPLEVRYTLIDKIGKRYNTQSCALLDNPRLNSGLPYALQFSG 465  
QY 344 -----KLEFYTERLVQHYNETLALSFSKGTLDONKRVNIYQDQERELIST-- 396  
Db 466 EGAVTISERWKERLA-PHFLQIQAI-PS-GKMHSQAHVFF-----PKLSGKLIAQDNEI 517  
QY 397 -----RGLSE-----QILTF--LFRLPKEAVLPICIKL----- 422  
Db 518 FIKGFRASBEKMPSSSLVYGTLSLPDLVSPFAFLQMTRYSPSLHSDGKALIK 577  
QY 423 -----LASQQLATTAISFLSHTSQHOALDLFQAA---KLPEPIIR- 463  
Db 578 GNVKLFITDPESPFLIEFKILIPDIVISSLDPSAPMTADNISVQASGELLQLPVDRILRL 637  
QY 464 AYADIAL-----YNLTKDPEKRSLSHDYAKKLIQETLL-----PVDTE-----NQR 504  
Db 638 QHKDLSISRYIGETSEASFOLYSPKDEETV-DISSRFKTDALTGDFRVNKNELSLTEK 696

OY 505 PHSWMLRYOYTYPE-----SRTKMLDLENTLATS--EDIRL 544  
 DB 697 THGS---LWMEISPERYSSFEKASCPCSCILHPTTRLDLSCLDKRGYISCLSL 753  
 OY 545 IOLMTEGDAKNPVL 559  
 DB 754 SOGIEGKLSSTPLV 768  
 RESULT 9  
 ID O9V455 PRELIMINARY; PRT; 514 AA.  
 AC O9V455;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE KARYOPHERIN-ALPHA3 PROTEIN.  
 GN KARYOPHERIN-ALPHA-3 OR CG9423.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Adill J.F., Agbayani A., An H.-J., Andrews-Piankocch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 Burtis K.C., Busan D.A., Butler H., Caden E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jaitai M., Kalush F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Mason D.A., Goldfarb D.S.;  
 RT "Drosophila melanogaster karyopherin alpha 3";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A.  
 DR Dockendorff T.C., Tang Z., Jongsens T.A.;

RT "Interaction cloning and characterization of karyopherin alpha 3 from  
 RT Drosophila";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003683; AAF54408.1; -  
 DR EMBL: AF152928; AAD37442.1; -  
 DR HSSP: 002821; 1BK5.  
 DR FLYBase: FBgn0027338; karyopherin-alpha-3.  
 DR InterPro: IPR000225; Armadillo.  
 DR InterPro: IPR002652; IBH.  
 DR Pfam: PF00514; Armadillo\_seg; 8.  
 DR Pfam: PF01749; IBH; 1.  
 DR SMART: SM00185; ARM; 6.  
 DR PROSITE: PS50176; ARM\_REPEAT; 2.  
 DR SEQUENCE 514 AA; 56990 MW; 0F500AC4EA90683a CRC64;  
 SQ  
 Query Match 4.7%; Score 132.5; DB 5; Length 514;  
 Best Local Similarity 21.2%; Pred. No. 0.96; Indels 137; Gaps 19;  
 Matches 98; Conservative 64; Mismatches 164;  
 OY 95 SEALDVLISQAMETAD--PLQDLVL SAVSGHIGKTSDDLFLKALASPPVIREAAVRL 151  
 DB 62 SSSIDLKRIAKAAADATKPEQOLA VQA-----ARKLL 94  
 OY 152 ANLKRKVIYDHLHSFTHLPEIQCLSAIFLRTEESDAIYRDLAAKSAISATRL 211  
 DB 95 SSDKNPPIINDLIQSDI--LPILVECLKQHNHTMLOFEAAAML-----TNLSGTS 143  
 OY 212 QIGEVQOKRFLPTLNLTLSASPODEALVALGK-----LKGQSYVNNKQLO--- 261  
 DB 144 QTNVYVAGAVPLFLQILNSAPNVCQAVNALGNIIGDGLIRFVIRKGVQPLSFI 203  
 OY 262 KPDVDVTLA--AAQALALGKEEDALPVYIKQALEERRALVALRH----- 305  
 DB 204 KPDPIITFLRVNTVYVNLCKRKNDA--PTATHEILPALNVLNHRDTNLYVTWAI 261  
 OY 306 -----LPSEIG-IFLALPIFLKTNSEAKLV-----ALALLEGCDT----- 342  
 DB 262 SYLTDGNDQIOMYIESGVKPLPL--GNSEVKVOTALRAVGNIVTSGDEQTVL 318  
 OY 343 -----PKLEYTERLVOPHNELTALSFSGKRTLONNKRVNIIVP----- 383  
 DB 319 NYDALSTPFGLSHPKEKIRK-----AVNFLSNITAGNSQVAVNVLGPKRIENL 372  
 OY 384 -----QDPERERLIS--TRGLEQILTFLEPKREAVLPCYIKLASQTOATTAIS 436  
 DB 373 SKGEQGTQKEAAMASINTISGNEQVFTL---KEGVIRPFDDLSCDPTQVINVVL 428  
 OY 437 FLSH-----TSHQAL-DLFLQAKLGEPTIRAVADLAITNL 473  
 DB 429 GLNNMLKVAADSHVAVANCICEGCLAKIERLOSHENVEIYKL 471  
 RESULT 10  
 ID O9UIK7 PRELIMINARY; PRT; 948 AA.  
 AC O9UIK7;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE ELKS.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Nakata T.;  
 RT "Fusion of a Novel Gene, ELKS, to c-ret in a Papillary Thyroid  
 RT Carcinoma";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB015617; BAA8763.1; -

DR InterPro: IPR002017; Spectrin.  
SQ SEQUENCE 948 AA; 108792 MW; 344297FDFC9F7602 CRC64;

Query Match 4.7%; Score 132.5; DB 4; Length 948;  
Best Local Similarity 20.1%; Pred. No. 2.3;  
Matches 116; Conservative 106; Mismatches 204; Indels 151; Gaps 28;

```

OY 35 ISTOSTQOALATYLEADGDHDFVLRKIGEDYLKQSIHSSDP-----QTRKSTIIIGA 89
DB 414 LSTEEEREE-----MKOMEYVRSHSKFNKNKIGQ--VKOESRKTELLAQTKLETLLNQ 467
OY 90 GLAGSSALDVLSCAMETADPLOQLVLVSAVSGHIGKTSDDLFLKALASPPYVRLBAAY 149
DB 468 -FDSQKHIEVLEKESLAKK--QRAAIL-----QTEVDAL-----RLRLKEKE 507
OY 150 RLANKTKYKID-----HLHSF-----THKLPEIQCISAAT----- 181
DB 508 TMLNKKTKROIODMAEEKTOAGEIHDLKMDLVKEREKVVNLQKKIENLOEQLRDEROMS 567
OY 182 -----FLRLTEESDAYTRDL---LAAKSATRSATALQIGEYQOKR-----FLPTLR 226
DB 568 SLKERVKSLOADDTNTPTALTTEELAEKERTERLEKQRDEREREKQOEIYDVKKDLK 627
OY 227 NL-----LTSASPDQDEAILYALGKLKDGOSYVNIKKOLQKPDVDTLAAQAALIALGKE 281
DB 628 DLKRVKSLDGLDSEKASLDDLKELHASSLASSGLKDSRLKTELEALQ-----KK 679
OY 282 EDALPV---IKK---QALEER--PRALYALRHLPSEIGIPALPIFLTKNSEAKLNVAL 333
DB 680 EECLEKMSOLKKAHEALAEARASPEMSDRIOHLEREI-----TRYKDESSKAQAEV-- 730
OY 334 ALLELGDTPKLELYTERLVPQHPYNE-----TLALSFSKGRTLQNMKRVNIIVPODQ 387
DB 731 -----DLLEITLKE--VENENKDKDKKIALELESITSQVQDNKKANLKHKEVE 779
OY 388 ERE--RLLSRTTGLEQILFLFLPKREAYLPCYKL---LASQKOTALATTAISFLSHTS 442
DB 780 KKKSAQMLEARREDNINDSSQOLQVEELLMAEMEKVQOELSMKAKLSIQOSLAERET 839
OY 443 H-----QEAIDLFOAKKLPREPITRAY-----ADLAIYNLTKDEPKKRSLDYAKKILQ 493
DB 840 HLNLRERKRKHLEVELEMKOEALLAISSEKDNAILLELSS-----SKKTKQE 888
OY 494 TILFVDTEQNRHPSMPYLRYQVTPRESRTKMLDILE 530
DB 889 EVALKREKDR--LVQOLKQOT--QNMKMLADNYE 920

RESULT 11
O9HKN6 PRELIMINARY; PRT; 1200 AA.
AC O9HKN6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN TA0561.
GN TA0561.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RA MEDLINE=20479972; PubMed=11029001;
RA Repp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum ";
RL Nature 407:508-513(2000).
DR EMBL; AL45064; CAC11701.1; -.
DR InterPro; IPR000547; Clathrin_repeat.

```

DR InterPro: IPR001623; DnaJ\_N.  
DR InterPro: IPR001440; TPR.  
DR Pfam: PF00515; TPR; 4.  
DR SMART: SM00289; CLH; 1.  
DR PROSITE: PS00636; DnaJ\_1; UNKNOWN\_1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1200 AA; 138346 MW; B056E2030B945418 CRC64;

Query Match 4.7%; Score 132.5; DB 17; Length 1200;  
Best Local Similarity 19.8%; Pred. No. 3.3;  
Matches 115; Conservative 88; Mismatches 209; Indels 169; Gaps 25;

```

OY 98 LDVLSQAMETADP-----LQQLVLVSAVSGHIGKTSDDLFLKALASPPYVRLBAAY 149
DB 99 LDVI--REVTSDPYFCEVENIVNLVISA--DHLEKYTFVAFILSKGCRVD--DRIAAYYA 153
OY 150 -RLANKTKYVIDHLHSFIHKLPEIQCISAATFLR--LTSASDAYTRDLA----- 199
DB 154 ERDRDDSTTDII-----IDNYEKHNDICRAVISFLRHVLAYEYSERYITKLECGRYED 207
OY 200 -----AKKSATRSATALQIGEYQOKRFLPTLRNLTLSASPDQEA-- 240
DB 208 SESVIYAVELAEPRVENADSSAILARTALDLGKYEKAQIAE--RGLKLNPDSEDLKLM 266
OY 241 ---LYALGKLKDGGSY-----NIKKOLQK 262
DB 267 ARSLYALGRINESLDYFQVDCVNIHPEKNREANYEMMDILYNGRKTEYATETLNVHRESFR 326
OY 263 PD-----VDVTLAAQAALIALGKEEDALPVIKKQALEERP-----RALYALRH 305
DB 327 PEDYIRMSGLIKDSGVDGSVNLLEALAKPPNDIIDIKAYAAAKKEIGNVGEALQAVON 386
OY 306 L-----PSEIGIPALPIFLTKNSEAKLNVAL-----LALTEIGCC--DT 342
DB 387 LIRIKPDDESLKFPVMDYFFGRNEDILNIYESLQDADLKEKYRGMALASVYIMGYMDA 446
OY 343 PKLLEYITERLVOPHYNETLALSFSKGRTLQNMKRVNIIVPODERERLSTTR--GLE 400
DB 447 VSMTRDREHLLDDPFVDSILFSVAKK-----EVQWMLISGLENYTSRLSVDFLFGIE 500
OY 401 ---EQILFELR-----LPKEAYLPCYKLLASQKOTALATTAISFLSHTSQOELD 448
DB 501 IRGVQYIIDYATKSCSKAMAEVAAEAVFRKNHTVADKIKVALSTKCL-----EYYD 552
OY 449 LLFQAAKLPGERPIIRAVADLAIYNLTG--DPEKRSRHHYAKKILQETILFVDTEQNRH 507
DB 553 IMISMGIYSKENVKIFEDHPRYLYPAIDTFTAGLYDEAYKILR-----YEGTKEDP 606
OY 508 SMPY---LRQVTPRESRTKMLDILETLATSKSSEDIRLL 544
DB 607 FLDIYEYARLNYM---NRTKDAIKILIRAKESFNSIDRYLL 644

RESULT 12
O30981 PRELIMINARY; PRT; 2554 AA.
AC O30981;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE FENGYCIN SYNTHETASE FENE.
GN FENE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F29-3;
RA Chen I., Lin G., Shu H., Liu S.;
RA "Analysis of the fenGYCIN Synthetase Gene fene ";
RT Thesis (1997), Microbiology and Immunology, Chang-Gung University.

```

DR EMBL: AF023465; AAB80956.1; -  
 DR HSSP: P14687; 1AMU.  
 DR InterPro: IPR000873; AMP-bind.  
 DR InterPro: IPR001242; DUF4.  
 DR InterPro: IPR003880; Phosphopant\_attach.  
 DR Pfam: PF00501; AMP-binding; 2.  
 DR Pfam: PF00668; Condensation; 3.  
 DR Pfam: PF00550; PP-binding; 2.  
 DR PRINTS: PR00154; AMPBINDING.  
 DR PROSITE: PS00075; ACP\_DOMAIN; 2.  
 DR PROSITE: PS00455; AMP\_BINDING; 2.  
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN\_1.  
 DR Phosphopantetheine.  
 DR SEQUENCE 2554 AA; 286724 MM; AEF0E71BF77AC0 CRC64;

Query Match 4.7%; Score 132.5; DB 2; Length 2554;  
 Best Local Similarity 20.0%; Pred. No. 9.7;  
 Matches 105; Conservative 80; Mismatches 186; Indels 155; Gaps 21;

QY 122 GHLGTSDDLFLKALASPPVIRLEAAYRLANKTKVIDHLHSFTIKLPEEIQCLSAAI 181  
 DB 1851 GYLGRPD---LTKKEFVNPFAPGEQMYRTGDLARW-LPDGTIEYGVDDQVKIRG--- 1903  
 QY 182 FLRLTEESDAYIRDLAALKSAIRSATALQIGEVQOKRFLPTLNLTLSASPODEAIL 241  
 DB 1904 -YRVLGETESALRHIDGKKEAAVLARTG-QLGTKE-----L 1938  
 QY 242 VALGKDGOSYYNIKLOKPDVDVTLAAAGALATLGEEDALPV---IRKQALEBR 296  
 DB 1939 YAVISVKEGTDAEQVRYTHLSQ-----MLPGYMPAYVIEMLDLTLNGKLNKRALEP- 1991  
 QY 297 PRALYALNHLSEIGIPALPIFLTKNSEAKLNVALLELGCPTPKLE----- 347  
 DB 1992 PDITSKQTYVPVRNDLEOLAIWQELVLTQIGIEDSFEELGSDISIKALQVARSALRGY 2051  
 QY 348 ---YTERLVOPHYNETLALSFSGRT-----LQNMKNRNIIVPOPORE 390  
 DB 2052 WSLHNSDLFRHAKIKDLSAIVKTERVAVIDQSGVAGVFWPTLPQHM-----FLSDQIKERH 2106  
 QY 391 R-----LSTGRGLEEQILTFELRLPKRAY--LPCIKYKLASOKTO-----LATTA 434  
 DB 2107 HFNQSVMLFSPDCLSENMLRASLKLAEHHDALRMVIREDSGQGMQINODIHESLYSLR 2166  
 QY 435 ISFLS-----HTSHQELDLFLQAKLPGEPIIRA-----YADLAIVNLTKD--- 476  
 DB 2167 ISDLSDSGMDWETSIKEEVANLQOSINLQGPPLHAAMFKTLGSDYLFLLTHHILVVDGVS 2226  
 QY 477 -----PEKKRSLMDYAK-----KLIOETLLFVDTENOR 504  
 DB 2227 WRILLEDSAAVHOASQAQALQLPKPTDSTYOETARRVODVYAOSSSKLIREETWYMSVEEEK 2286  
 QY 505 ---PHSPMYLRYQVTPESRTKLMLDLLETPLATSKSSESDIRLLIQ 546  
 DB 2287 AAEPLPY-EMPYMENMNSSER-----ETLRPSLTADAVALLQ 2322

RESULT 13  
 ID 044295 PRELIMINARY; PRT; 398 AA.  
 AC 044295;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE ORF398 PROTEIN.  
 GN ORF398.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103650;  
 RN [1]  
 RC STRAIN=PCC 7120;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97144534; PubMed=8990301;

RA Frlas J.E., Flores E., Herrero A.;  
 RT "Nitrate assimilation gene cluster from the heterocyst-forming  
 RT cyanobacterium Anabaena sp. Strain PCC 7120.";  
 RL J. Bacteriol. 179:477-486(1997).  
 DR EMBL: X99708; CAA68039.1; -  
 DR InterPro: IPR004155; HEAT\_PBS.  
 DR Pfam: PF03130; HEAT\_9.  
 DR SEQUENCE 398 AA; 43548 MM; A38946D4ID5AF05E CRC64;

Query Match 4.6%; Score 131; DB 2; Length 398;  
 Best Local Similarity 22.6%; Pred. No. 0.84;  
 Matches 90; Conservative 38; Mismatches 123; Indels 148; Gaps 12;

QY 31 KILYISTOSTOQALATYLEALDAYGDHDFVLRKIGEDYLKOSIHSSDPQTRKSTLIGAG 90  
 DB 65 KVLVMGNVTINPLDILDEDEDAEDLRWFARIRIGE----- 101  
 QY 91 LAGSSEALDVLQSAMETADPLQOOLVLSAVSGHLCKTSDDLFLKALAS-----PPY 141  
 DB 102 -----LQHPBEALPVELL-----KTSDEEIKAIASSAIAQMGTLAIP 140  
 QY 142 VI-----RLAEAYRLANKTKVIDHLHSFTIKLPEEIQCLSAAILFLRLTEESDA 192  
 DB 141 VIVELLAQENTRLAVRSIAYIRHTQTAPLSVQ-----DTQAS-- 181  
 QY 193 YIRDLAALKSAIRSATALQIGEVQOKRFLPTLNLTLSASPODEAILVALGKLDGOS 252  
 DB 182 -----IRAAAEIALSFDQVNPFLNALNDLSTVTRTAIQGSFSPDLS 229  
 QY 253 YINIKQLOKPR---DVDTYLAAGALATLGEEDALVYIKQALEERP-----RAL 300  
 DB 230 ELNLVAQIQPKLYDENIECCAAANALAMGDDAAOHLYKLISANHPITLQLEIRAL 289  
 QY 301 YALRHLPs-----EIGIPALPIFLTKNS----- 325  
 DB 230 VMLESLISLEVIQCAFNOVTTETLMQELVTVYVGRQKRELPK-ATAILLELNSPHRAT 348  
 QY 326 ---EAKLNVALALTELG---CDPRKLEYYTERLVQPH 357  
 DB 349 KHSKSAIALSLGOLGNAAEATESLTMLSTPDELAVRH 387

RESULT 14  
 ID 09EP71 PRELIMINARY; PRT; 979 AA.  
 AC 09EP71;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE NORPEG-LIKE PROTEIN (ANKYCORBIN).  
 GN 1700020L1IRK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC STRAIN=C57 BL/6J;  
 RA Peng Y.-F., Mandai K., Sakisaka T., Okabe N., Yamamoto Y.,  
 RA Yokoyama S., Mizoguchi W., Shiozaki H., Monden M., Takai Y.;  
 RT "Ankycorbin: a novel actin cytoskeleton-associated protein.";  
 RL Genes Cells 0:0-0(2000).  
 DR EMBL: AF274866; AAG25937.1; -  
 DR EMBL: AF202315; AAG24483.1; -  
 DR HSSP: P42773; 1IHB.

DR MGD, MG1:1922896; 1700020L1IR1K.  
 DR InterPro; IPR002110; ANK.  
 DR PRINTS; PRO1415; ANKYRIN.  
 DR SMART; SM00248; ANK; 6.  
 DR PROSITE; PSS0088; ANK\_REPEAT; 5.  
 DR PROSITE; PSS0297; ANK\_REGION; 1.  
 DR ANK repeat; Repeat.  
 KW SEQUENCE 979 AA; 108851 MW; B2B8C016D80237C4 CRC64;

Query Match 4.6%; Score 130.5; DB 11; Length 979;  
 Best Local Similarity 20.5%; Pred. No. 3.3;  
 Matches 127; Conservative 89; Mismatches 224; Indels 181; Gaps 28;

QY 32 ILVTSQSTQALATVLEA-----LDAVGHDFVLRKIGEDYVKOSIHS 76  
 DB 336 LLDISSADODLLVLAQAVASITLHNKELQDKLQAKSKPD-----KEARADISFOSFHS 391  
 QY 77 SD-----PQTRKSTIIGAGLAGSSEALD-----VLSQAMETADPLQQL----- 115  
 DB 392 TQTDLAPSPGKASDIPSSDAKSSPVEHPAGTSTTDNDVIRIROLQDSLHDLQKRLSESEA 451  
 QY 116 -----VLSVSGHLGKTSDDL--LFKALASPY-----PVIRLEAAYPLAN 153  
 DB 452 EKKQLODELQSORDTLCLINTEISENGSDLSQKLKETOSEKYEAMKEVLSVOKOMKIGL 511  
 QY 154 LKNTKVIDHLSFTHKPEEIOCSAIFRLTEES--DAYTRD--LAAKSAIRS 207  
 DB 512 LSGESADGYSHLKEAPADEIDTLKQD--LQKAVEESARKKERELETLAEQAEAT 569  
 QY 208 ATALQIGEYQOKREPLRLNI-----LTSASPODEAILYALGKLKDGOSYVNIKKOL 260  
 DB 570 KPPAEACEELRSYSYIENNNKKAFLFEKYQQAQEIIM-----TKLD-----TLKSQMP 620  
 QY 261 -QKPD--VDYTLAAQALILGKEEDALPYI-----KKQALEERPRALYALRH 305  
 DB 621 QEAPDSDGDKEMNNRMLDELNKOVSSELSQLYREAOALEDDYRKRSLEDAAEYIHKAH 680  
 QY 306 LPSEIGIPALFELKTRN--SEAKLNVALLLELGCCTPKLEYITE--RLVQPHYMETL 362  
 DB 681 -----ERLMHVSNI-SRAKSEALS--EKKSOYSKVLNLTQLKLVDAH----- 722  
 QY 363 ALSFSKRTLQNMKNRNIIVPODERELISTTGLLEQILFLFRLP--KEAVLPCTYK 421  
 DB 723 -----KENSVESTENLQVITTLRTAKEMEKISALTGLHAKAEAVALEK 769  
 QY 422 LLSQKQLATTAISFLSHTSHOEALDLFOAAKLPEPIIRAYADLAITYNLTKDPEKR 481  
 DB 770 QLAEEKAASDAMVPKSSYEKLQASL-----ESEVNALA-----TKLKSVR 811  
 QY 482 SLHDYAKKLQOETLLFVDENORHPSPMYLRQVTPESRTKMLDLLETLATSKSEDI 541  
 DB 812 -----BREKASHVAQVRSVSQAREK--DNIOQLLAKKEQD-V 848  
 QY 542 RLIIQLTMEGDANKFVYAGL 562  
 DB 849 TALVQKFORAOE-----LAGM 865

RESULT 15  
 Q9M6P9 PRELIMINARY; PRT; 1316 AA.  
 AC Q9M6P9;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE DNA REPAIR-RECOMBINATION PROTEIN.  
 GN RAD50.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN-CV, COLUMBIA.  
 RX MEDLINE=21097002; Pubmed=11169180;  
 RA Gallego M.E., Jeanneau M., Granier F., Bouchez D., Bechtold N.,  
 RA White C.I.;  
 RT "Disruption of the Arabidopsis RAD50 gene leads to plant sterility and  
 RT RMS sensitivity";  
 RL Plant J. 25:31-41(2001).  
 DR EMBL; AF168748; AAF36810.1; -  
 DR InterPro; IPR003439; ABC\_transport.  
 DR InterPro; IPR001687; ATP\_GTP\_A.  
 DR SEQUENCE 1316 AA; 152814 MW; 89DC4F6BCA39B0E8 CRC64;

Query Match 4.6%; Score 129.5; DB 10; Length 1316;  
 Best Local Similarity 22.0%; Pred. No. 5.9;  
 Matches 127; Conservative 93; Mismatches 229; Indels 129; Gaps 25;

QY 62 LRKIGEDYVKOSI-----HSSDPQTRKSTI-----IGAGLAGSEALDVLQSOA 104  
 DB 745 LRAVFEEXSKLTITITLAEKTLQEHTEELGQSEALDVLGISAQIKADKDSIEALVQ 804  
 QY 105 METADPLQQLLVLSAVSGHLGKTSDDLFRALASPYVIRLEAY--RLANKTKVIDHL 163  
 DB 805 LENADRIFOEIV-----SYQKQIEDLEYKIDFRGLGVTKMEEIOSELSIQSSK--DKL 856  
 QY 164 HSFTHKPE-----EIOCSAIF--LRLETESDAYIRLLAKSAIRSATALQGE 215  
 DB 857 HGELEKLRDQIYMERISCIQARWHAVEREKAANILRLVTAEDLERLA----- 909  
 QY 216 YQOKRFLPTLRNLITSA--SPQDEAIL--YALGKLKDGOSYNI--KKQLOKPDVDT 268  
 DB 910 -EEKSQDLDVQKYLTEALPLSKREKQLDSYNNMKIRRNQYVELAKKRNQOQVEAL 968  
 QY 269 LAAQALIA--IGKEDALPVIKQALEERPRALYALRHLPSEIGIPALPIFLKTRNS 325  
 DB 969 LKASYKINEYHDLKGERLIDQEKORLSDS-----QLQSC 1004  
 QY 326 EAKLNVALL--ELGCDTPKLEYITTELQVPHYNETLALSFSGRTLQNMKR-- 377  
 DB 1005 EARNELAGELNRNKLDRNODQLRNIENL--NYRTKAKVEELRELESLEQILN 1061  
 QY 378 -----VNIIVPODERERLLS-----TTTGLEQILFLRLPREAYLPCTYKLA 424  
 DB 1062 IGGIAVAEAEIVKILREBERILSELNCRGTVSYESSISKNRVE--KQAOYKIDIRHF 1120  
 QY 425 SOKTOLATTAISFLSHTSHOEALD--LFOAAKLPG--EPIIRAY-----ADLAITYNL 473  
 DB 1121 DQLIQLTTEWANKDDRYVNALDKALRFHTMMEELIKIIRELMQOTYRGQMDYIRI 1180  
 QY 474 TKDPEKRSLSHDYAKKLQOETLLFVDENORHPSPMYLRQVTPESRTKMLDLLETLA 533  
 DB 1181 HSDSE--GAGTRSYSKVLMQV--GDTE-----LEWRGCSAGOKVLASLIIRLA 1226  
 QY 534 TSKS--SEDIRLIQLTMEGDANKFVYAGLIIKIVE 568  
 DB 1227 LAETFLNCGILLALDEPTTINDGPNSESLAGALLRIME 1264

Search completed: August 20, 2002, 04:23:45  
 Job time: 809 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:19:11 ; Search time 17.02 Seconds  
(without alignments)  
815.143 Million cell updates/sec

Title: US-09-662-812-2  
Perfect score: 568  
Sequence: 1 MGLPRLTFLGLCSLPISL.....TEGDANKPPVLAGLILIKIVE 568

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/pcodata/2/1aa/5A.COMB.pep:\*  
3: /cgn2\_6/pcodata/2/1aa/5B.COMB.pep:\*  
4: /cgn2\_6/pcodata/2/1aa/6A.COMB.pep:\*  
5: /cgn2\_6/pcodata/2/1aa/6B.COMB.pep:\*  
6: /cgn2\_6/pcodata/2/1aa/6CTUS.COMB.pep:\*  
7: /cgn2\_6/pcodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.2	11	US-07-657-769B-43	Sequence 43, App1
2	7	1.2	11	US-07-789-184-104	Sequence 104, App
3	7	1.2	11	US-07-789-184-215	Sequence 215, App
4	7	1.2	11	US-07-789-184-218	Sequence 218, App
5	7	1.2	11	US-08-475-263-104	Sequence 104, App
6	7	1.2	11	US-08-475-263-215	Sequence 215, App
7	7	1.2	11	US-08-475-263-218	Sequence 218, App
8	7	1.2	11	US-08-485-886-104	Sequence 104, App
9	7	1.2	11	US-08-485-886-215	Sequence 215, App
10	7	1.2	11	US-08-485-886-218	Sequence 218, App
11	7	1.2	11	US-08-477-362-104	Sequence 104, App
12	7	1.2	11	US-08-477-362-215	Sequence 215, App
13	7	1.2	11	US-08-477-362-218	Sequence 218, App
14	7	1.2	11	US-08-477-134-104	Sequence 104, App
15	7	1.2	11	US-08-477-134-215	Sequence 215, App
16	7	1.2	11	US-08-477-134-218	Sequence 218, App
17	7	1.2	11	US-08-473-489A-104	Sequence 104, App
18	7	1.2	11	US-08-473-489A-215	Sequence 215, App
19	7	1.2	11	US-08-473-489A-218	Sequence 218, App
20	7	1.2	11	US-08-485-695-104	Sequence 104, App
21	7	1.2	11	US-08-485-695-215	Sequence 215, App
22	7	1.2	11	US-08-485-695-218	Sequence 218, App
23	7	1.2	11	US-08-018-760-104	Sequence 104, App
24	7	1.2	11	US-08-018-760-215	Sequence 215, App
25	7	1.2	11	US-08-018-760-218	Sequence 218, App
26	7	1.2	151	US-08-905-223-436	Sequence 436, App
27	7	1.2	186	US-08-936-165A-534	Sequence 534, App

28	7	1.2	226	3	US-08-651-136C-16	Sequence 16, App1
29	7	1.2	247	1	US-08-446-083-5	Sequence 5, App1
30	7	1.2	293	3	US-08-651-136C-20	Sequence 20, App1
31	7	1.2	298	3	US-08-651-136C-18	Sequence 18, App1
32	7	1.2	322	1	US-08-118-270-36	Sequence 36, App1
33	7	1.2	322	5	PCIT-US93-08528-36	Sequence 36, App1
34	7	1.2	331	4	US-08-849-751-4	Sequence 4, App1
35	7	1.2	331	4	US-09-413-231-4	Sequence 4, App1
36	7	1.2	331	4	US-09-478-816-4	Sequence 4, App1
37	7	1.2	335	2	US-08-505-218-4	Sequence 4, App1
38	7	1.2	367	4	US-08-888-429A-13	Sequence 13, App1
39	7	1.2	425	1	US-07-657-769B-69	Sequence 69, App1
40	7	1.2	425	1	US-08-097-938-7	Sequence 7, App1
41	7	1.2	425	1	US-08-313-553-13	Sequence 13, App1
42	7	1.2	425	1	US-07-789-184-220	Sequence 220, App
43	7	1.2	425	1	US-08-476-000-7	Sequence 7, App1
44	7	1.2	425	1	US-08-475-263-220	Sequence 220, App
45	7	1.2	425	1	US-08-472-840-7	Sequence 7, App1

ALIGNMENTS

RESULT 1  
US-07-657-769B-43  
; Sequence 43, Application US/07657769B  
; Patent No. 5256766  
; GENERAL INFORMATION:  
; APPLICANT: COUGHLIN, SHAUN R.  
; TITLE OF INVENTION: RECOMBINANT RECEPTOR AND RELATED  
; NUMBER OF INVENTIONS: PHARMACEUTICALS  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: IRELL & MANELLA  
; STREET: 545 MIDDLEFIELD ROAD, SUITE 200  
; CITY: MENLO PARK  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/657,769B  
; FILING DATE: 19910219  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2000-0502.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-327-7250  
; TELEFAX: 415-327-2951  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-07-657-769B-43

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 436 SFLSHTS 442  
Db 3 SFLSHTS 9

*Issued  
Patents*

RESULT 2  
US-07-789-184-104  
Sequence 104, Application US/07789184  
Patent No. 5688768  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
ATTORNEY/AGENT INFORMATION:  
SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/789,184  
FILING DATE: 19911107  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-789-184-104

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442  
|||||||  
DB 3 SFLSHTS 9

RESULT 3  
US-07-789-184-215  
Sequence 215, Application US/07789184  
Patent No. 5688768  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
ATTORNEY/AGENT INFORMATION:  
SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/789,184  
FILING DATE: 19911107  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 218:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-789-184-218

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/789,184  
FILING DATE: 19911107  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 215:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-789-184-215

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442  
|||||||  
DB 3 SFLSHTS 9

RESULT 4  
US-07-789-184-218  
Sequence 218, Application US/07789184  
Patent No. 5688768  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
ATTORNEY/AGENT INFORMATION:  
SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/789,184  
FILING DATE: 19911107  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 218:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-789-184-218

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442  
DB 3 SFLSHTS 9

RESULT 5  
US-08-475-263-104  
; Sequence 104, Application US/08475263  
; Patent No. 5759994  
; GENERAL INFORMATION:  
; APPLICANT: COUGHLIN, SHAUN R.  
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
; NUMBER OF SEQUENCES: 223  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,263  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELETYPE: 90-4030  
; INFORMATION FOR SEQ. ID NO: 104:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-475-263-104

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442  
DB 3 SFLSHTS 9

RESULT 6  
US-08-475-263-215  
; Sequence 215, Application US/08475263  
; Patent No. 5759994  
; GENERAL INFORMATION:  
; APPLICANT: COUGHLIN, SHAUN R.  
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
; NUMBER OF SEQUENCES: 223

CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,263  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 22000-20502.03  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELETYPE: 90-4030  
; INFORMATION FOR SEQ. ID NO: 215:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-475-263-215

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 SFLSHTS 442  
DB 3 SFLSHTS 9

RESULT 7  
US-08-475-263-218  
; Sequence 218, Application US/08475263  
; Patent No. 5759994  
; GENERAL INFORMATION:  
; APPLICANT: COUGHLIN, SHAUN R.  
; TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
; NUMBER OF SEQUENCES: 223  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/475,263  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 22000-20502.03  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 218:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-475-263-218

Query Match 1.2%: Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442  
|||||||  
Db 3 SFLSHTS 9

RESULT 8  
US-08-485-886-104  
Sequence 104, Application US/08485886  
Patent No. 5798248  
GENERAL INFORMATION:

APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
TITLE OF INVENTION: RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,886  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/789,184  
FILING DATE: 07-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154

INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-485-886-104

Query Match 1.2%: Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442  
|||||||  
Db 3 SFLSHTS 9

RESULT 10  
US-08-485-886-218  
Sequence 218, Application US/08485886  
Patent No. 5798248  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
TITLE OF INVENTION: RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California

Db 3 SFLSHTS 9

RESULT 9  
US-08-485-886-215  
Sequence 215, Application US/08485886  
Patent No. 5798248  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
TITLE OF INVENTION: RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,886  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/789,184  
FILING DATE: 07-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154

INFORMATION FOR SEQ ID NO: 215:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-485-886-215

Query Match 1.2%: Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442  
|||||||  
Db 3 SFLSHTS 9

RESULT 10  
US-08-485-886-218  
Sequence 218, Application US/08485886  
Patent No. 5798248  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
TITLE OF INVENTION: RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California

Query Match 1.2%: Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442  
|||||||  
Db 3 SFLSHTS 9

RESULT 10  
US-08-485-886-218  
Sequence 218, Application US/08485886  
Patent No. 5798248  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
TITLE OF INVENTION: RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California

Query Match 1.2%: Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442  
|||||||  
Db 3 SFLSHTS 9

RESULT 10  
US-08-485-886-218  
Sequence 218, Application US/08485886  
Patent No. 5798248  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
TITLE OF INVENTION: RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California

COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,886  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/789,184  
FILING DATE: 07-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 218:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-485-886-218

Query Match 1.2%; Score 7; DB 1; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 436 SFLSHTS 442  
|||||  
Db 3 SFLSHTS 9

RESULT 11  
US-08-477-362-104  
Sequence 104, Application US/08477362  
Patent No. 5849507  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,362  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/789,184  
FILING DATE: 07-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-477-362-104

Query Match 1.2%; Score 7; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 436 SFLSHTS 442  
|||||  
Db 3 SFLSHTS 9

RESULT 12  
US-08-477-362-215  
Sequence 215, Application US/08477362  
Patent No. 5849507  
GENERAL INFORMATION:  
APPLICANT: COUGHLIN, SHAUN R.  
APPLICANT: SCARBOROUGH, ROBERT M.  
TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND  
RELATED PHARMACEUTICALS  
NUMBER OF SEQUENCES: 223  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,362  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/789,184  
FILING DATE: 07-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20502.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 34-0154  
INFORMATION FOR SEQ ID NO: 215:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-477-362-215

Query Match 1.2%; Score 7; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 436 SFLSHTS 442

Db 3 SFLSHTS 9

RESULT 13

US-08-477-362-218  
Sequence 218, Application US/08477362

Patent No. 5849507

GENERAL INFORMATION:

APPLICANT: COUGHLIN, SHAWN R.

APPLICANT: SCARBOROUGH, ROBERT M.

TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND

TITLE OF INVENTION: RELATED PHARMACEUTICALS

NUMBER OF SEQUENCES: 223

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,362

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/789,184

FILING DATE: 07-NOV-1991

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 22000-20502.20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

TELEX: 34-0154

INFORMATION FOR SEQ ID NO: 218:

SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-477-362-218

Query Match 1.2%; Score 7; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442

Db 3 SFLSHTS 9

RESULT 14

US-08-477-134-104  
Sequence 104, Application US/08477134

Patent No. 5856448

GENERAL INFORMATION:

APPLICANT: COUGHLIN, SHAWN R.

APPLICANT: SCARBOROUGH, ROBERT M.

TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND

TITLE OF INVENTION: RELATED PHARMACEUTICALS

NUMBER OF SEQUENCES: 223

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,134

FILING DATE: 07-JUN-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/789,184

FILING DATE: 07-NOV-1991

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 22000-20502.20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

TELEX: 34-0154

INFORMATION FOR SEQ ID NO: 104:

SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-477-134-104

Query Match 1.2%; Score 7; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SFLSHTS 442

Db 3 SFLSHTS 9

RESULT 15

US-08-477-134-215  
Sequence 215, Application US/08477134

Patent No. 5856448

GENERAL INFORMATION:

APPLICANT: COUGHLIN, SHAWN R.

APPLICANT: SCARBOROUGH, ROBERT M.

TITLE OF INVENTION: RECOMBINANT THROMBIN RECEPTOR AND

TITLE OF INVENTION: RELATED PHARMACEUTICALS

NUMBER OF SEQUENCES: 223

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,134

FILING DATE: 07-JUN-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/789,184

FILING DATE: 07-NOV-1991

ATTORNEY/AGENT INFORMATION:

NAME: MURASHIGE, KATE H.

REGISTRATION NUMBER: 29,959

REFERENCE/DOCKET NUMBER: 22000-20502.20  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 813-5600  
 TELEFAX: (415) 494-0792  
 TELEX: 34-0154  
 INFORMATION FOR SEQ ID NO: 215:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 11 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-477-134-215

Query Match 1.2%; Score 7; DB 2; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 436 SPLSHTS 442  
 |||||  
 Db 3 SPLSHTS 9

Search completed: August 20, 2002, 04:25:56  
 Job time: 405 sec

THIS PAGE BLANK (USPTO)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 01:09:15 ; Search time 1588.14 Seconds  
(without alignments)  
16206.816 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907

Sequence: 1 gtgctgtgatttgaaaaag.....ctcatcaagagagctgtgaaa 1907

Scoring table:

Gapop 60.0, Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estcp1:\*  
7: em\_estcp2:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.1	453	11 AK019607	AK019607 Mus muscu
2	21	1.1	458	9 AW098474	AW098474 ga03g06.x
3	21	1.1	534	9 AW680975	AW680975 WS1_9_H08
4	21	1.1	582	9 AW564895	AW564895 LG1_312_D
5	21	1.1	909	12 A2538423	A2538423 ENTFT26TF
6	20	1.0	216	10 BG959111	BG959111 PM4-CT080
7	20	1.0	224	9 AA669611	AA669611 ac19e05.s
8	20	1.0	326	10 BB649446	BB649446 UI-R-BH2.
9	20	1.0	370	9 AV855875	AV855875 AV855875
10	20	1.0	393	12 AQ124704	AQ124704 HS_2266.A
11	20	1.0	433	9 A1159433	A1159433 v279b05.f
12	20	1.0	452	10 BM409407	BM409407 EST583734
13	20	1.0	470	10 BJ073983	BJ073983 BJ073983
14	20	1.0	491	10 B1729956	B1729956 603350062
15	20	1.0	496	10 BM409927	BM409927 EST584254
16	20	1.0	510	10 BG098375	BG098375 EST462894
17	20	1.0	563	10 BE436245	BE436245 EST407323

C 18	20	1.0	567	10 BE433437	BE433437 EST399966
C 19	20	1.0	581	12 AQ276990	AQ276990 CITB1-E1-
C 20	20	1.0	590	9 AV852345	AV852345 AV852345
C 21	20	1.0	593	9 AV680699	AV680699 AV680699
C 22	20	1.0	607	9 A1486589	A1486589 EST244910
C 23	20	1.0	621	10 B1730573	B1730573 603350953
C 24	20	1.0	645	10 BM387094	BM387094 UI-R-CN1-
C 25	20	1.0	660	9 A1898005	A1898005 EST267448
C 26	20	1.0	673	9 BB649694	BB649694 BB649694
C 27	20	1.0	687	10 B1789662	B1789662 UI-R-DK0-
C 28	20	1.0	705	9 A1898469	A1898469 EST267912
C 29	20	1.0	713	10 BM411178	BM411178 EST585505
C 30	20	1.0	714	10 BG967518	BG967518 602834090
C 31	20	1.0	744	10 BM412698	BM412698 EST587025
C 32	20	1.0	759	9 AU080658	AU080658 AU080658
C 33	20	1.0	815	10 BG975641	BG975641 602845351
C 34	20	1.0	835	10 BF784222	BF784222 602108086
C 35	20	1.0	963	12 CNS03MW3	CNS03MW3 Tetradon
C 36	20	1.0	972	10 BG295852	BG295852 602933359
C 37	19	1.0	101	9 BE076598	BE076598 CM1-BT059
C 38	19	1.0	133	9 AA706483	AA706483 ag95d11.s
C 39	19	1.0	190	9 BB391458	BB391458 BB391458
C 40	19	1.0	326	9 AA09747	AA09747 OK88B09.s
C 41	19	1.0	350	9 AA961192	AA961192 Co52e01.s
C 42	19	1.0	357	9 A1451953	A1451953 mp75g06.x
C 43	19	1.0	373	10 BG930973	BG930973 f33-1134
C 44	19	1.0	380	10 BF018004	BF018004 uw96b04.x
C 45	19	1.0	380	10 BF018281	BF018281 uw97a10.x

## ALIGNMENTS

RESULT 1	AK019607	443 bp	mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930441N18:unclassified, full insert sequence.
LOCUS	AK019607		
DEFINITION	AK019607		
ACCESSION	AK019607.1	GI:12859904	
VERSION	AK019607		
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (sites)		
JOURNAL	Carninci, P., and Hayashizaki, Y.		
MEDLINE	High-efficiency full-length cDNA cloning		
PUBMED	Meth. Enzymol. 303, 19-44 (1999)		
AUTHORS	2 (sites)		
JOURNAL	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)		
AUTHORS	3 (sites)		
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hatada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, N., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
MEDLINE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)		
AUTHORS	20530913		

PUBMED 11076861  
 4 (sites)  
 REFERENCE  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FUNCTIONAL ANNOTATION  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE 5 (bases 1 to 443)  
 AUTHORS Adachi, J., Alizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,  
 Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C.,  
 Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,  
 Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,  
 Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,  
 Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S.,  
 Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,  
 Numata, K., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,  
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
 Schirral, L., Shibata, K., Shibata, Y., Shinagawa, A., Shinkai, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamana, I.,  
 Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and  
 Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,  
 URL: http://genome.gscc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216)  
 COMMENT Please visit our web site (http://genome.gscc.riken.go.jp/) for  
 further details.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..443  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="MGI:1912078"  
 /db\_xref="taxon:10090"  
 /clone="1930441N18"  
 /sex="male"  
 /issue\_type="testis"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 1..443  
 /note="evidence:NAS  
 unclassifiable"  
 misc\_feature  
 143 a 103 c 98 g 99 t  
 BASE COUNT  
 ORIGIN  
 Query Match 1.18; Score 21; DB 11; Length 443;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 299 gaagactatctcaagcaaac 319  
 |||||||||||||||||||||  
 131 GAAGACTATCTCAAGCAAGC 151

RESULT 2  
 AM098474/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 ORGANISM  
 SOURCE  
 Ceratodon purpureus.  
 Ceratodon purpureus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;  
 Bryopsida; Dicranidae; Dicranales; Ditracheaceae; Ceratodon.  
 1 (bases 1 to 458)  
 QUATRONO, R., BASHLARD, S., COVE, D., CUMING, A., KNIGHT, C., CLIFTON  
 S., MARRA, M., HILLIER, L., PAGE, D., MARTIN, J., WYLIE, T., UNDERWOOD  
 K., THEISING, B., ALLEN, M., BOWERS, Y., PERSON, B., SWALLER, T.,  
 STEPTOE, M., GIBBONS, M., HARVEY, N., RITTER, E., JACKSON, Y., MCCANN, R.,  
 WATERSTON, R. and WILSON, R.  
 Leeds/Mash U Moss EST Project  
 Unpublished (1999)  
 Other-ESTs: ga03906.y1  
 Contact: Ralph Quatran  
 Leeds/Mash U Moss EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 Libraries were constructed by Dr. Stavros Bashlades as part of the  
 Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and  
 Washington Univ. in St. Louis (USA) DNA sequencing by: Washington  
 University Genome Sequencing Center for information on obtaining a  
 clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 449.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..458  
 /organism="Ceratodon purpureus"  
 /db\_xref="taxon:3225"  
 /clone="PEP\_SOURCE\_ID:CPU011211"  
 /clone\_lib="Moss EST library CPU"  
 /tissue\_type="protonemata: 7 day old tissue"  
 /lab\_host="DH10B"  
 /note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2:  
 XhoI; Construction of the cDNA library was carried out  
 using Stratagene's 'UniZAP - cDNA synthesis kit'. cDNA was  
 constructed using an oligo dt primer/linker that contains  
 a XhoI site within it. Following ds cDNA synthesis, EcoRI  
 adapters were ligated to the blunt ends and sample was  
 digested with XhoI. The result is cDNA with an EcoRI  
 sticky end on one side and a XhoI sticky end on the other.  
 This cDNA was ligated directionally in UniZAP arms. The  
 vector is designed containing the bluescript sequence as  
 well as lambda DNA and cDNA is cloned within this  
 bluescript sequence. The vector was then packaged using  
 Gold gigaPackaging extracts. Library was grown in XLBlue  
 MRF cells and amplified. The library was excised by mass  
 excision using Stratagene's 'Mass excision kit' that uses  
 exsist as a helper phage that releases the bluescript  
 sequence and circularises it as single stranded plasmids  
 that are then packaged (by helper phage) and secreted out  
 of the host cell as phagemids. SOLR cells were transformed  
 with phagemids and the library was plated out on LB-amp  
 plates to select for transformants. Approximately 1,000  
 ,000 colonies were grown and recovered. The double  
 stranded plasmid library was recovered by using Qiagen  
 Midi prep kit. 2 micro grams of each library were used to  
 transform DH10B cells by electroporation."  
 141 a 51 c 121 g 144 t 1 others  
 BASE COUNT  
 ORIGIN  
 Query Match 1.18; Score 21; DB 9; Length 458;

Best Local Similarity 100.0%; Pred. No. 12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 583 tcatctacattcttcttica 603  
|||||  
Db 350 TCATCTACATCTTCTTCA 330

## RESULT 3

LOCUS AW680975 534 bp mRNA linear EST 19-JUL-2000  
DEFINITION WS1\_9\_H08.bl\_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA, mRNA  
sequence.

ACCESSION AW680975  
VERSION AW680975.1 GI:7554901  
KEYWORDS EST.

SOURCE sorghum.  
ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 534)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt  
L.H.

TITLE An EST database from Sorghum: water-stressed plants  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@prattuga.edu

Sequences have been trimmed to exclude polyA, vector and regions  
below phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 531  
POLYA=NO.

## FEATURES

## source

1..534  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Water-stressed 1 (WS1)"  
/note="Organ: Mix of 5-week old plants on days 7 & 8 after  
water was withheld; Vector: Lambda Zap; Site\_1: XhoI;  
Site\_2: EcoRI; The library was made from poly-A RNA in the  
cloning vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 168 a 143 c 108 g 115 t  
ORIGIN

Query Match 1.1%; Score 21; DB 9; Length 534;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 gcatctccctacatctgtcattc 529  
|||||  
Db 84 GCATCTCCCTATCTCTGCATC 104

## RESULT 4

LOCUS AW564895 582 bp mRNA linear EST 19-JUL-2000  
DEFINITION LG1\_312\_D09.bl\_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA  
sequence.

ACCESSION AW564895  
VERSION AW564895.1 GI:7218773  
KEYWORDS EST.

SOURCE sorghum.  
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 582)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.  
TITLE An EST database from Sorghum: light-grown seedlings  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmp@prattuga.edu

Sequences have been trimmed to exclude polyA, vector and regions  
below phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 566  
POLYA=NO.

## FEATURES

## source

1..582  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Light Grown 1 (LG1)"  
/note="Organ: 10- to 14-day-old light-grown (greenhouse)  
seedlings; Vector: Lambda Zap; Site\_1: XhoI; Site\_2: EcoRI  
; The library was made from poly-A RNA in the cloning  
vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."  
BASE COUNT 174 a 159 c 135 g 114 t  
ORIGIN

Query Match 1.1%; Score 21; DB 9; Length 582;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 509 gcatctccctacatctgtcattc 529  
|||||  
Db 183 GCATCTCCCTATCTCTGCATC 203

## RESULT 5

LOCUS A2538423 909 bp DNA linear GSS 14-NOV-2000  
DEFINITION ENTFT26TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.

ACCESSION A2538423  
VERSION A2538423.1 GI:11143345  
KEYWORDS GSS.

SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.

REFERENCE Eukaryota; Entamoebidae; Entamoeba.  
1 (bases 1 to 909)  
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library

JOURNAL Unpublished (2000)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjl@loftuseligr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library

Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 18  
High quality sequence stop: 770.

location/Qualifiers

## FEATURES

source 1..909

/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: pHOSt1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described from Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaubin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 288 a 111 c 170 g 340 t  
ORIGIN

Query Match 1.1%; Score 21; DB 12; Length 909;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 556 taattgaagaactaagt 576  
|||||  
Db 875 TAATTGAAGAACTAAGT 895

RESULT 6  
BG959111 216 bp mRNA linear EST 12-JUN-2001  
DEFINITION PM4-CT0806-180301-003-d02 CT0806 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG959111  
VERSION BG959111.1 GI:14377282  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 216)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM4&ct=PM4-CT0806-  
180301-003-d02&ts=2001-03-18&ts=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 9  
High quality sequence stop: 216.

FEATURES  
SOURCE 1. 216  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0806"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
/716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 62 a 44 c 58 g 52 t  
ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 216;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 ttgaaagatcagatg 30  
|||||  
Db 168 TTTGAAAAGTCATGAT 187

RESULT 7  
AA669611 224 bp mRNA linear EST 20-NOV-1997  
LOCUS AA669611/c  
DEFINITION ac19e05.s1 Stratagene ovary (#937217) Homo sapiens cDNA clone  
IMAGE:856928 3', mRNA sequence.  
ACCESSION AA669611  
VERSION AA669611.1 GI:2631110  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 224)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin  
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,  
White,Y., Wylie,T., Waterston,R. and Wilson,R.  
WashU-NCI human EST Project  
Unpublished (1997)  
JOURNAL Contact: Wilson Rk  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40ml3 fwd. En from Amersham.

FEATURES  
SOURCE 1. 224  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:856928"  
/clone\_lib="Stratagene ovary (#937217)"  
/sex="Female"  
/dev\_stage="49 year old"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: ovary; Vector: Bluescript SK; Site\_1: EcoRI;  
Site\_2: XhoI; Cloned unidirectionally". Primer: Oligo dt.  
Total ovary tissue, normal, caucasian. Average insert  
size: 0.8 kb; Uni-Zap XR Vector; -5' adaptor sequence: 5'  
GAATTCGGACGAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTT 3"  
BASE COUNT 54 a 47 c 25 g 98 t  
ORIGIN

Query Match 1.0%; Score 20; DB 9; Length 224;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 64 catgaatacttctag 83  
|||||

Db 176 CATTGAATCTGCTAGAG 157

RESULT 8 326 bp mRNA linear EST 06-SEP-2000  
 BE649446  
 LOCUS  
 DEFINITION UI-M-BH2.2-son-b-10-0-UI.r1 NIH\_BMAP\_M.S3.2 Mus musculus cDNA clone  
 BE649446  
 ACCESSION UI-M-BH2.2-son-b-10-0-UI 5', mRNA sequence.  
 VERSION BE649446.1 GI:9975235  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Bernaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mestr@mail.nih.gov  
 cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It  
 should be noted that Bento Soares is generating a small number of  
 additional specialized non-redundant arrays of BMAP cDNAs whose  
 availability will be considered under appropriate and limited  
 collaborative arrangements  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1..326  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH2.2-son-b-10-0-UI"  
 /clone\_1id="NIH\_BMAP\_M.S3.2"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pMT3D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; The  
 NIH\_BMAP\_M.S3.2 library is a subtracted library of a  
 series, ultimately derived from a mixture of individually  
 tagged normalized libraries from ten regions of the mouse  
 brain (cerebellum, brain stems, olfactory bulbs,  
 hypothalamus, cortex, amygdala, basal ganglia, pineal  
 gland, striatum, hippocampus) after a series of  
 subtractions to reduce the representation of cDNAs from  
 which ESTs had already been generated. The following  
 serially subtracted libraries were generated in this  
 process: NIH\_BMAP\_M.S3.2, NIH\_BMAP\_M.S2, NIH\_BMAP\_M.S1.  
 The subtracted library (NIH\_BMAP\_M.S3.2) was constructed  
 as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M.S2  
 clones from which 3' ESTs had been derived was used as a  
 driver in a hybridization with the NIH\_BMAP\_M.S2 library  
 in the form of single-stranded circles. The remaining  
 single-stranded circles (subtracted library) was purified  
 by hydroxapatite column chromatography, converted to  
 double-stranded circles and electroporated into DH10B  
 bacteria (Life Technologies) to generate the  
 NIH\_BMAP\_M.S3.2 library. This procedure has been  
 previously described (Bernaldo, Lennon and Soares, Genome  
 Research 6:791-806, 1996)"

BASE COUNT 99 a 72 c 82 g 73 t

Query Match 1.0%; Score 20; DB 10; Length 326;

Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1423 ctccatcaggaagccttag 1442  
 ||||||||||||||||  
 Db 77 CTCACATCAGGAGACCTTAG 96

RESULT 9 370 bp mRNA linear EST 08-NOV-2001  
 AV855875/c  
 LOCUS  
 DEFINITION AV855875 Nori Satoh unpublished cDNA library, egg Clona  
 AV855875  
 ACCESSION AV855875  
 VERSION AV855875.1 GI:16843399  
 KEYWORDS EST.  
 SOURCE Clona intestinalis.  
 ORGANISM Clona intestinalis  
 Eukaryota; Metazoa; Chordata; Orochordata; Ascidiacea; Enterogona;  
 Phlebobranchia; Cloniidae; Clona.

REFERENCE  
 AUTHORS Satoh,N., Satou,Y., Kohara,Y. and Shin-i,T.  
 TITLE Expressed genes in Clona intestinalis  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Nori Satoh  
 Department of Zoology  
 Kyoto University  
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
 Tel: 81-75-753-4081  
 Fax: 81-75-705-1113  
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.  
 Location/Qualifiers  
 1..370  
 /organism="Clona intestinalis"  
 /db\_xref="taxon:7719"  
 /clone="rcieg22b12"  
 /clone\_1id="Nori Satoh unpublished cDNA library, egg"  
 /tissue\_type="whole animal"  
 /dev\_stage="egg"

BASE COUNT 127 a 48 c 62 g 130 t 3 others

ORIGIN

Query Match 1.0%; Score 20; DB 9; Length 370;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 124 ctggagcttattgtgta 143  
 ||||||||||||||||  
 Db 57 CTTGGACTTTATGTGCTA 38

RESULT 10 393 bp DNA linear GSS 22-SEP-1998  
 A0124704/c  
 LOCUS  
 DEFINITION HS-2266.AL.D10.MF CIT Approved Human Genomic Sperm Library D Homo  
 sapiens genomic clone Plate=2266 Col=19 Row=G, DNA sequence.  
 A0124704  
 ACCESSION A0124704.1 GI:3501870  
 VERSION  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1332657"
/clone_lib="Soares_mammary_gland_NbMG"
/sex="male"

```

prior to freezing the pericarp." 82 c 117 g 120 t

# ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 agctctagtcataagtc 196  
 ||||||||||||||||  
 Db 245 AGTCTGTAGTCATAGATC 226

## RESULT 13

LOCUS B1073983 470 bp mRNA linear EST 11-DEC-2001  
 DEFINITION B1073983 N1BB Mochii normalized Xenopus tailbud library Xenopus  
 laevis cDNA clone XL110103 5', mRNA sequence.

ACCESSION B1073983  
 VERSION B1073983.1 GI:17504172  
 KEYWORDS EST.  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 470)  
 AUTHORS Kiteyama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara Y.  
 TITLE Expressed genes in X. laevis embryo  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Tadasu Shin-I  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@genes.nig.ac.jp.  
 Location/Qualifiers

## FEATURES

source 1..470  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="XL110103"  
 /clone\_lib="N1BB Mochii normalized Xenopus tailbud library"  
 /tissue\_type="whole embryo"  
 /dev\_stage="stage 25"  
 BASE COUNT 168 a 75 c 122 g 104 t 1 others  
 ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 929 gcttggggaagaagaaga 948  
 ||||||||||||||||  
 Db 435 GCTTGGGGAAGAAGAGA 454

## RESULT 14

LOCUS B1729956 491 bp mRNA linear EST 20-SEP-2001  
 DEFINITION 603350062F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:5357703 5',  
 mRNA sequence.

ACCESSION B1729956  
 VERSION B1729956.1 GI:15706982  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 491)  
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
 Tissue Procurement: The Cepko Laboratory  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM11909 row: k column: 16  
 High quality sequence stop: 488.

FEATURES 1..491  
 Location/Qualifiers

source 1..491  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5357703"  
 /clone\_lib="NIH\_MGC\_94"  
 /tissue\_type="retina"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: eye; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 3.3 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."  
 BASE COUNT 140 a 104 c 144 g 103 t  
 ORIGIN

Query Match 1.0%; Score 20; DB 10; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1423 ctccatcaggaagccttag 1442  
 ||||||||||||||||  
 Db 289 CTCACATCAGGAAGCCTTAG 308

## RESULT 15

LOCUS BM409927/c 496 bp mRNA linear EST 22-JAN-2002  
 DEFINITION EST584254 tomato breaker fruit Lycopersicon esculentum cDNA clone  
 CLEG50D8 5' end, mRNA sequence.

ACCESSION BM409927  
 VERSION BM409927.1 GI:18261557  
 KEYWORDS EST.  
 SOURCE tomato.  
 ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 496)  
 AUTHORS Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai J., Bougri,O., Kirkness,E., Uterback,T., Van Aken,S., Rinning C.M., Fraser,C.M., Martin,G.B., Tanksey,S.D. and Giovannoni,J.  
 Generation of ESTs from tomato fruit tissue, breaker stage (2002)  
 Unpublished (2002)  
 Contact: CUGI  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 This clone is available through the Clemson University Genomics  
 Institute  
 Seq primer: T3.

FEATURES 1..496  
 Location/Qualifiers  
 source 1..496  
 /organism="Lycopersicon esculentum"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="CLEG50D8"

```
/clone_lib="tomato breaker fruit"
/classue_type="Pericarp"
/dev_stage="Breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCuaapt; Site.1: EcoRI;
Site.2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."
BASE COUNT      149 a      88 c      127 g      132 t
ORIGIN
```

```
Query Match      1.0%; Score 20; DB 10; Length 496;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 177 agctctgttagtcatagaatc 196
      |||||||||||||||||||
Db 245 AGTCTGTAGTGCATATAGATC 226
```

Search completed: August 20, 2002, 02:30:56  
Job time: 4901 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 01:14:20 ; Search time 55.18 Seconds  
(without alignments)  
8489.001 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907  
Sequence: 1 gtgagcttgattgaaaaaag.....ctcatcaagagcgtgtgaa 1907

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/1na/PCTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	0.9	401	US-09-221-298-45	Sequence 45, Appl
2	18	0.9	1030	US-08-858-207A-102	Sequence 102, App
3	18	0.9	10091	US-09-058-489-34	Sequence 34, Appl
4	17	0.9	454	US-08-680-326-118	Sequence 118, App
5	17	0.9	957	US-08-858-207A-104	Sequence 104, App
6	17	0.9	2057	US-09-082-089-4	Sequence 4, Appl
7	17	0.9	2190	US-08-082-089-1	Sequence 1, Appl
8	17	0.9	2625	US-08-759-945-1	Sequence 1, Appl
9	16	0.8	57	US-08-888-381-8	Sequence 8, Appl
10	16	0.8	84	US-08-182-175A-27	Sequence 27, Appl
11	16	0.8	84	US-08-182-175A-28	Sequence 28, Appl
12	16	0.8	84	US-08-182-175A-28	Sequence 28, Appl
13	16	0.8	84	US-08-474-633A-68	Sequence 68, Appl
14	16	0.8	84	US-08-474-633A-69	Sequence 69, Appl
15	16	0.8	84	PCT-US92-06412-27	Sequence 27, Appl
16	16	0.8	84	PCT-US92-06412-28	Sequence 28, Appl
17	16	0.8	175	US-08-182-175A-58	Sequence 58, Appl
18	16	0.8	175	US-08-474-633A-76	Sequence 76, Appl
19	16	0.8	175	PCT-US92-06412-58	Sequence 58, Appl
20	16	0.8	273	US-09-050-159-126	Sequence 126, App
21	16	0.8	399	US-08-894-699-18	Sequence 18, Appl
22	16	0.8	399	US-09-444-410-18	Sequence 18, Appl
23	16	0.8	681	US-08-329-681A-14	Sequence 14, Appl
24	16	0.8	782	US-08-261-825-1	Sequence 1, Appl
25	16	0.8	782	US-08-719-124-1	Sequence 1, Appl
26	16	0.8	782	PCT-US95-07748-1	Sequence 1, Appl
27	16	0.8	930	PCT-US95-07748A-1	Sequence 1, Appl
				US-08-936-165A-74	Sequence 74, Appl

C	28	16	0.8	1028	1	US-08-029-328-1	Sequence 1, Appl
	29	16	0.8	1400	2	US-08-481-658B-43	Sequence 43, Appl
	30	16	0.8	1400	2	US-08-477-504A-43	Sequence 43, Appl
	31	16	0.8	1400	2	US-08-486-756A-43	Sequence 43, Appl
	32	16	0.8	1400	2	US-08-485-862B-43	Sequence 43, Appl
	33	16	0.8	1400	3	US-08-787-739-43	Sequence 43, Appl
	34	16	0.8	1400	3	US-08-487-077A-43	Sequence 43, Appl
	35	16	0.8	1400	3	US-08-485-863A-43	Sequence 43, Appl
	36	16	0.8	1400	4	US-08-485-049D-43	Sequence 43, Appl
	37	16	0.8	1400	4	US-09-178-115-43	Sequence 43, Appl
	38	16	0.8	1400	4	US-09-177-776-43	Sequence 43, Appl
	39	16	0.8	1480	1	US-08-484-105-19	Sequence 19, Appl
	40	16	0.8	1480	1	US-08-484-106-19	Sequence 19, Appl
	41	16	0.8	1496	4	US-09-384-212-1	Sequence 1, Appl
	42	16	0.8	1552	3	US-08-948-705-1	Sequence 1, Appl
	43	16	0.8	1691	3	US-08-948-564-3	Sequence 3, Appl
	44	16	0.8	1692	4	US-08-858-207A-39	Sequence 39, Appl
	45	16	0.8	1789	3	US-08-948-705-6	Sequence 6, Appl

## ALIGNMENTS

```

RESULT 1
US-09-221-298-45
: Sequence 45, Application US/09221298
: Patent No. 6284241
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiaqichun
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY AND DIAGNOSIS
: FILE REFERENCE: 210121.471
: CURRENT APPLICATION NUMBER: US/09/221,298
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 45
: LENGTH: 401
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (212)
: OTHER INFORMATION: Where n is a, c, g or t
: NAME/KEY: modified_base
: LOCATION: (224)
: OTHER INFORMATION: Where n is a, c, g or t
: US-09-221-298-45

Query Match 0.9%; Score 18; DB 4; Length 401;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1749 aagagatgcaaaaatt 1766
Db 152 aagagatgcaaaaatt 169

RESULT 2
US-08-858-207A-102/c
: Sequence 102, Application US/08858207A
: Patent No. 6348328
: GENERAL INFORMATION:
: APPLICANT: Black, Michael
: APPLICANT: Hodgson, John
: APPLICANT: Knowles, David
: APPLICANT: Nicholas, Richard
: APPLICANT: Stodola, Robert
: TITLE OF INVENTION: No. 6348328el Compounds
: NUMBER OF SEQUENCES: 552
: CORRESPONDENCE ADDRESS:

```

ADDRESS: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,207A  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/017670  
FILING DATE: 14-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50475  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 102:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1030 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-858-207A-102

Query Match 0.9%; Score 18; DB 4; Length 1030;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 aaactctgacgactta 493  
|||||  
Db 369 AAAACTCTGACGACTTA 352

RESULT 3  
US-09-058-489-34  
Sequence 34, Application US/09058489  
Patent No. 6103886  
GENERAL INFORMATION:  
APPLICANT: Whitehead Institute for Biomedical Research  
APPLICANT: Lahn, Bruce  
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of  
FILE REFERENCE: WH197-089A  
CURRENT APPLICATION NUMBER: US/09/058,489  
CURRENT FILING DATE: 1998-04-10  
EARLIER APPLICATION NUMBER: 60/041,877  
EARLIER FILING DATE: 1997-04-11  
NUMBER OF SEQ ID NOS: 91  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 34  
LENGTH: 10091  
TYPE: DNA  
ORGANISM: Human  
US-09-058-489-34

Query Match 0.9%; Score 18; DB 3; Length 10091;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1439 ttatagctacttccaa 1456

Db 8676 ttatagctacttccaa 8693  
|||||

RESULT 4  
US-08-680-326-118  
Sequence 118, Application US/08680326  
Patent No. 5925733  
GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARINX  
APPLICANT: STRAND, KURT  
APPLICANT: TODARO, GEORGE J.  
TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES  
TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL  
NUMBER OF SEQUENCES: 152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/680,326  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Schliff, J. Michael  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 454 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..454  
US-08-680-326-118

Query Match 0.9%; Score 17; DB 2; Length 454;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 476 aaactctgacgactt 492  
|||||  
Db 139 AAAACTCTGACGACTT 155

RESULT 5  
US-08-858-207A-104/c  
Sequence 104, Application US/08858207A  
Patent No. 6348328  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Nicholas, Richard  
APPLICANT: Stodola, Robert  
TITLE OF INVENTION: No. 6348328el Compounds

NUMBER OF SEQUENCES: 552  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,207A  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/017670  
FILING DATE: 14-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50475  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 957 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-858-207A-104

Query Match 0.9%; Score 17; DB 4; Length 957;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 tgcacagcttgcaga 737  
Db 423 TGCACAGCTTGCAGA 407

RESULT 6  
US-09-082-089-4/c  
Sequence 4, Application US/09082089  
Patent No. 6100060  
GENERAL INFORMATION:  
APPLICANT: BARNES, MICHAEL  
APPLICANT: TESTA, TANIA  
APPLICANT: KELSEL, DAVID  
TITLE OF INVENTION: No. 6100060e1 Compounds  
NUMBER OF SEQUENCES: 5  
CURRENT APPLICATION DATA:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/082,089  
FILING DATE: 20-MAY-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9710737.9  
FILING DATE: 23-MAY-1997  
APPLICATION NUMBER: GB 9803981.1  
FILING DATE: 25-FEB-1998  
APPLICATION NUMBER: GB 9804007.4  
FILING DATE: 25-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F.  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-30166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701  
TELEX: 846169  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2057 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-082-089-4

Query Match 0.9%; Score 17; DB 3; Length 2057;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 907 agcagcagctcaagctt 923  
Db 1924 AGCAGCAGCTCAAGCTT 1908

RESULT 7  
US-09-082-089-1/c  
Sequence 1, Application US/09082089  
Patent No. 6100060  
GENERAL INFORMATION:  
APPLICANT: BARNES, MICHAEL  
APPLICANT: TESTA, TANIA  
APPLICANT: KELSEL, DAVID  
TITLE OF INVENTION: No. 6100060e1 Compounds  
NUMBER OF SEQUENCES: 5  
CURRENT APPLICATION DATA:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RATNER & PRESTIA  
STREET: P.O. BOX 980  
CITY: VALLEY FORGE  
STATE: PA  
COUNTRY: USA  
ZIP: 19482  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/082,089  
FILING DATE: 20-MAY-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9710737.9  
FILING DATE: 23-MAY-1997  
APPLICATION NUMBER: GB 9803981.1  
FILING DATE: 25-FEB-1998  
APPLICATION NUMBER: GB 9804007.4  
FILING DATE: 25-FEB-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESTIA, PAUL F.  
REGISTRATION NUMBER: 23,031  
REFERENCE/DOCKET NUMBER: GH-30166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-407-0700  
TELEFAX: 610-407-0701

TELEX: 846169  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2190 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-09-082-089-1

Query Match 0.9%; Score 17; DB 3; Length 2190;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 907 agcagcagctcaagctt 923  
|||||  
Db 2057 AGCAGCAGCTCAAGCTT 2041

RESULT 8  
US-08-759-945-1  
; Sequence 1, Application US/08759945  
; Patent No. 5834249  
; GENERAL INFORMATION:  
; APPLICANT: KAZUAKI, Furukawa  
; APPLICANT: SUGIMURA, Keijiro  
; APPLICANT: OHSUKE, Kazuhiro  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: 699 Prince Street  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/759, 945  
; FILING DATE: 03-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/303,191  
; FILING DATE: 08-SEP-1994  
; APPLICATION NUMBER: JP 5-257881  
; FILING DATE: 08-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rea, Teresa Stanek  
; REGISTRATION NUMBER: 30,427  
; REFERENCE/DOCKET NUMBER: 001560-227  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2625 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..2625  
; NAME/KEY: mat\_peptide  
; LOCATION: 118..2625  
US-08-759-945-1

Query Match 0.9%; Score 17; DB 2; Length 2625;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 467 catcttggaacttc 483  
|||||  
Db 2427 CATCTTGGAACCTTC 2443

RESULT 9  
US-08-888-381-8  
; Sequence 8, Application US/08888381  
; Patent No. 6110703  
; GENERAL INFORMATION:  
; APPLICANT: Egel-Mitani, Michi  
; APPLICANT: Brandt, Jakob  
; APPLICANT: Vad, Knud  
; TITLE OF INVENTION: Method For Production of Polypeptides  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6110703 No. 6110703disk of No. 6110703th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/888,381  
; FILING DATE: 04-June-1999  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gregg, Valeta A.  
; REGISTRATION NUMBER: 35,127  
; REFERENCE/DOCKET NUMBER: 4840.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 57 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-888-381-8

Query Match 0.8%; Score 16; DB 3; Length 57;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1353 gatttataaagctttt 1368  
|||||  
Db 33 GTATTATAAGCTTTT 48

RESULT 10  
US-08-182-175A-27/C  
; Sequence 27, Application US/08182175A  
; Patent No. 5559223  
; GENERAL INFORMATION:  
; APPLICANT: Saverio Carl Falco  
; APPLICANT: Sharon J. Keeler  
; APPLICANT: Janet A. Rice  
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESS:

```

: ADDRESSEE: E.I. du Pont de Nemours and Company
: STREET: 1007 Market Street
: CITY: Wilmington
: STATE: Delaware
: COUNTRY: USA
: ZIP: 19898
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: OPERATING SYSTEM: Macintosh System, 6.0
: SOFTWARE: Microsoft Word, 4.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/182,175A
: FILING DATE:
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/743,006
: FILING DATE: 9 August 1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Linda Axamethy Floyd
: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: BB-1031
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (302) 992-4929
: TELEFAX: (302) 892-7949
: TELEX: 835420
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 84 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..84
: OTHER INFORMATION: /product="synthetic oligonucleotide"
: OTHER INFORMATION: /standard_name="SM 98"
: US-08-182-175A-27

Query Match      0.8%; Score 16; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagcctgagttctc 1200
   |||||||
Db 44 TAGCCTTGAGTTCTC 29

RESULT 11
US-08-182-175A-28
: Sequence 28, Application US/08182175A
: Patent No. 5559223
: GENERAL INFORMATION:
: APPLICANT: Saverio Carl Falco
: APPLICANT: Sharon J. Keeler
: TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
: NUMBER OF SEQUENCES: 113
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E.I. du Pont de Nemours and Company
: STREET: 1007 Market Street
: CITY: Wilmington
: STATE: Delaware
: COUNTRY: USA
: ZIP: 19898
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: OPERATING SYSTEM: Macintosh System, 6.0
: SOFTWARE: Microsoft Word, 4.0
: CURRENT APPLICATION DATA:
```

```

: APPLICATION NUMBER: US/08/182,175A
: FILING DATE:
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/743,006
: FILING DATE: 9 August 1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Linda Axamethy Floyd
: REGISTRATION NUMBER: 33,692
: REFERENCE/DOCKET NUMBER: BB-1031
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (302) 992-4929
: TELEFAX: (302) 892-7949
: TELEX: 835420
: INFORMATION FOR SEQ ID NO: 28:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 84 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1..84
: OTHER INFORMATION: /product="synthetic oligonucleotide"
: OTHER INFORMATION: /standard_name="SM 99"
: US-08-182-175A-28

Query Match      0.8%; Score 16; DB 1; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagcctgagttctc 1200
   |||||||
Db 44 TAGCCTTGAGTTCTC 59

RESULT 12
US-08-474-633A-68/C
: Sequence 68, Application US/08474633A
: Patent No. 5773691
: GENERAL INFORMATION:
: APPLICANT: E. I. DU PONT DE NEMOURS AND
: APPLICANT: COMPANY
: TITLE OF INVENTION: CHIMERIC GENES AND
: TITLE OF INVENTION: METHODS FOR INCREASING
: TITLE OF INVENTION: INCREASING THE LYSINE
: TITLE OF INVENTION: AND THREONINE CONTENT
: TITLE OF INVENTION: OF THE SEEDS OF PLANTS
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: E. I. DU PONT DE NEMOURS
: ADDRESS: AND COMPANY
: STREET: 1007 MARKET STREET
: CITY: WILMINGTON
: STATE: DELAWARE
: COUNTRY: U.S.A.
: ZIP: 19898
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: MICROSOFT WORD VERSION 2.0C
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,633A
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: BARBARA C. SIEGELL
: REGISTRATION NUMBER: 30,684
: REFERENCE/DOCKET NUMBER: BB-1037-C
: TELECOMMUNICATION INFORMATION:
```

TELEPHONE: 302-992-4931  
TELEFAX: 302-773-0164  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..84  
OTHER INFORMATION: /product= "synthetic"  
OTHER INFORMATION: /oligonucleotide= "SM"  
OTHER INFORMATION: /standard\_name= "SM"  
US-08-474-633A-68

Query Match 0.8%; Score 16; DB 1; Length 84;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1185 tagcctgagttctc 1200  
|||||  
Db 44 TAGCCTGAGTTCTC 29

RESULT 13  
US-08-474-633A-69  
Sequence 69, Application US/08474633A  
Patent No. 5773691  
GENERAL INFORMATION:  
APPLICANT: E. I. DU PONT DE NEMOURS AND  
APPLICANT: COMPANY  
TITLE OF INVENTION: CHIMERIC GENES AND  
TITLE OF INVENTION: METHODS FOR INCREASING  
TITLE OF INVENTION: INCREASING THE LYSINE  
TITLE OF INVENTION: AND THREONINE CONTENT  
TITLE OF INVENTION: OF THE SEEDS OF PLANTS  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. DU PONT DE NEMOURS  
ADDRESS: AND COMPANY  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: U.S.A.  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD VERSION 2.0C  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,633A  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: BARBARA C. SIEGEL  
REGISTRATION NUMBER: 30,684  
REFERENCE/DOCKET NUMBER: BB-1037-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4931  
TELEFAX: 302-773-0164  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 69:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..84  
OTHER INFORMATION: /product= "synthetic"  
OTHER INFORMATION: /oligonucleotide= "SM"  
OTHER INFORMATION: /standard\_name= "SM"  
US-08-474-633A-69

Query Match 0.8%; Score 16; DB 1; Length 84;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1185 tagcctgagttctc 1200  
|||||  
Db 44 TAGCCTGAGTTCTC 59

RESULT 14  
PCT-US92-06412-27/C  
Sequence 27, Application PC/TUS9206412  
GENERAL INFORMATION:  
APPLICANT: Saverio Carl Falco  
APPLICANT: Sharon J. Keeler  
APPLICANT: Janet A. Rice  
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E.I. du Pont de Nemours and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: USA  
ZIP: 19898  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06412  
FILING DATE: 19920807  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/743,006  
FILING DATE: 9 August 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Linda Axamethy Floyd  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: BB-1031  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302) 992-4929  
TELEFAX: (302) 892-7949  
TELEX: 835420  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..84  
OTHER INFORMATION: /product= "synthetic oligonucleotide"  
OTHER INFORMATION: /standard\_name= "SM 98"  
PCT-US92-06412-27

Query Match 0.8%; Score 16; DB 5; Length 84;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagccttgagttctc 1200  
 |||  
 Db 44 TAGCCTTGAGTTCTC 29

RESULT 15  
 PCT-US92-06412-28  
 ; Sequence 28, Application PC/TUS9206412  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Saverio Carl Falco  
 ; APPLICANT: Sharon J. Keeler  
 ; APPLICANT: Janet A. Rice  
 ; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F  
 ; NUMBER OF SEQUENCES: 113  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: E.I. du Pont de Nemours and Company  
 ; STREET: 1007 Market Street  
 ; CITY: Wilmington  
 ; STATE: Delaware  
 ; COUNTRY: USA  
 ; ZIP: 19898  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disk  
 ; COMPUTER: Macintosh  
 ; OPERATING SYSTEM: Macintosh System, 6.0  
 ; SOFTWARE: Microsoft Word, 4.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/06412  
 ; FILING DATE: 19920807  
 ; CLASSIFICATION: 530  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/743,006  
 ; FILING DATE: 9 August 1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Linda Axamechy Floyd  
 ; REGISTRATION NUMBER: 33,692  
 ; REFERENCE/DOCKET NUMBER: BB-1031  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (302) 992-4929  
 ; TELEFAX: (302) 892-7949  
 ; TELEX: 835420  
 ; INFORMATION FOR SEQ ID NO: 28:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 84 base pairs  
 ; TYPE: NUCLEIC ACID  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1..84  
 ; OTHER INFORMATION: /product="synthetic oligonucleotide"  
 ; OTHER INFORMATION: /standard\_name="SM 99"  
 ; PCT-US92-06412-28

Query Match 0.8%; Score 16; DB 5; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1185 tagccttgagttctc 1200  
 |||  
 Db 44 TAGCCTTGAGTTCTC 59

Search completed: August 20, 2002, 03:12:04  
 Job time: 7064 sec

THIS PAGE BLANK (USPTO)



QY 1861 ttatgggttcctcagctgatgcttgaactcatcaagaagcgtgtgaaa 1907  
|||||  
Db 1861 TTATGGGTTCCAGCTGATGCTTTGACATCATCAAGAGCCTGTGAAA 1907  
1

RESULT 2  
LOCUS AE002234 8268 bp DNA linear BCN 30-MAY-2000  
DEFINITION Chlamydophila pneumoniae AR39, section 62 of 94 of the complete genome.  
ACCESSION AE002234 AE002161  
VERSION AE002234.2 GI:8163492  
KEYWORDS  
SOURCE Chlamydophila pneumoniae AR39.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 8268)  
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
PUBMED 10684935  
JOURNAL 2 (bases 1 to 8268)  
MEDLINE Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA  
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189667.  
FEATURES  
source 1. 8268  
/organism="Chlamydophila pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
113..1036  
/gene="CP0751"  
113..1036  
/gene="CP0751"  
/note="similar to GB:000096 SP:Q47012 PID:1381661  
PID:178603; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="AtpA/E1ac family protein"  
/protein\_id="AAF38556.1"  
/db\_xref="GI:7189668"  
/transl\_table=1  
/translation="MSSRELILGSSQOPTRTNCGAVLFRWNGEGLTFDEGEQTOR  
OFIPANIATQVNRIFVSHFHDGDCGLGSMRLRLNDKVSFHCYVPSAGSKYDR  
LRVGTIVHETIQVNEHPISSEGIQVDSFRIEAOQLQHOVDLIGRIRTEPQIKLP  
KELSRGKIGLIIQDILRQDEISIGSTYVLSVYRKGDSTAIATADLPQCAATDL  
ANSCMCEESTYLDQHRHLASHFHTAKQATLAKRATOKLILTHPSARYLNIDD  
FYKEASAVPPNSVAQEVRSYFPKPNPLNK"  
1121..2059  
/gene="CP0752"  
1121..2059  
/gene="CP0752"  
/note="CP0752"  
/note="integrase/recombinase, phage integrase family;  
identified by match to PFAM protein family HMM PF00589"  
/codon\_start=1  
/transl\_table=11  
/product="Integrase/recombinase, phage integrase family"  
/protein\_id="AAF73701.1"  
/db\_xref="GI:8163493"  
/transl\_table=1  
/translation="MTASIVSEFLDYLKWKVKSAPHTLRNYCIDLNLKIFLEGRNIA  
PSSPIQATATEKRVKSELPSLFTKEHVRMYIAKLINGRAKAKTIKRLSSISFAHYC

gene  
CDS  
2114..3706  
/gene="CP0753"  
2114..3706  
/gene="CP0753"  
/note="ABC transporter, ATP-binding protein; identified by  
match to PFAM protein family HMM PF00005"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, ATP-binding protein"  
/protein\_id="AAF73702.1"  
/db\_xref="GI:8163494"  
/transl\_table=1  
/translation="MSIVLDKIGKSGITFLDDVSVPENPCNGYGLTPNGAGKSTL  
LKTIMGIEPTKSGISLPKRGILRONIDSFDHTVLDGVKGNTRLMALQNRDLY  
LOEFTDAIGMEIGETIEIIGENGRYADSEAEELTGIGIPNMFPRKAMAPIDLOF  
RVLCQALGHEALIDEPYTNLDLYSINWLNFNLDYEGYIVVSHDRHPLNTTT  
HIAIDIDYITIPGNYDMVEKRTASROEKADIKSKKKISOLKEFYAKFGAGSRA  
SOVOSLRREIKLQPOELKKSNIQRPYRFPISDKSSGRVVISLAEITFDYGHQVHY  
PSLEYQGDKIGITGNGLGKTTMLKLAGVAPSSGSIKQHQAICSPYPNHSDY  
LADCGEITLFEWLRNRKTSINDQETRSVLGKMLFGGDGAFKQIALSGGETARLLMAG  
MMLENNHNVLIDENNNHLDLESVSALSMAINDYKTAIFVSHDRGLIDQCATKLIFD  
KDKTFEFDCTMVDYTAGHKOLL"  
3735..4325  
/gene="CP0754"  
3735..4325  
/gene="CP0754"  
/note="similar to GB:108793 SP:002169 GB:M96343 PID:142853  
PID:143161; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="maf protein"  
/protein\_id="AAF38557.1"  
/db\_xref="GI:7189669"  
/transl\_table=1  
/translation="MSLPVLGSSSPRRKFLIEKFAVPEFVIPSNFDESKVSYGDP  
AYTOELAAQKAVAVSELHSPDCIILTGPIYVSDGRIFKPDQKADAIOMLTKLRNO  
THDVYTSIVAIKRGKLLNGSESOISLTWMPDRISYIDYGTGLNCGAAYDCRGL  
ILKRVGCVYNOGLPIOTLKIILEELNDMDYST"  
4307..6013  
/gene="CP0755"  
4307..6013  
/gene="CP0755"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF38558.1"  
/db\_xref="GI:7189670"  
/transl\_table=1  
/translation="WGFHILTLFGLLCSPISLVAKFPESVGHKILYISTQOAL  
AYLEALDVAQGDHDFVLRKIGEDYILKOSIHSSDPQTRSTIIGAGLASSSEALDVIS  
QAMETADPLQOQLVLSAVSGHLGKTSDDLDFKALSPYPIRLAAYRUALNKNVYI  
DHLSHFIRKLPEBIOCLSAIFRLRTEESDAITRDLAAKSAISAVALQIGEQO  
KRLPLRILTPESPODEAILIYALGKLDGGSYNNIKKOLOKPOVDATLAAOQIL  
ALCKEEDTAPVKQKALLERPRALVLRHPISGIGIPALPIELKNSKAPKINVA  
LLEAGCDTPKLEFYITERLVQPIRYNETLALSFSGKRTLOMKKVVNIIVDDPDERRL  
LSTRLEQDILTEFLRLEAVLPCTIVKLASQKQALATTAISFLGSHOABDL  
FQAKLPGEPPIIRAVDLAIYNTKDPKKRSADYAKALIQETLLFVDENGRPHS  
MPALRYQVTPESRTKLMDILETLATSKSESDIRLLIQMTGEDAKNFPVLAGLLIKI  
VE"  
6003..8117  
/gene="CP0756"  
6003..8117  
/gene="CP0756"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAF38559.1"  
/db\_xref="GI:7189671"

/translaction="MMSNPILRLMKRCFLFASPVLMSGSSADALTTHOEAVKKNSTYS  
HFKSVSGITVIEDGVNLHNHNRIOANKVVENPVMSGSLTVAIGNMNVNRAATVLC  
DYLEYVDPDSCILNTRFAMYPFELGSMITLTPETIVIRKGIISTEGKNDLCS  
GDYLESSDLSLIGTKTILNCRIPILFLPPEISIMPEIRKPNFNGGJGFLSFL  
GMSYSPISIRHFSSTFELDFSKHGVNGMGNLHCSQOVEMENFMKSYAHLRAIIM  
AEDHRYRLHGDCEFTHKHVNSEGYHLSDSMVEVADIPFNMFILKMGPRVCCWN  
DNEGYLTSSKVNPNONOEPLYLQOYSPINVTYENIYECGYLNAPSD  
HYGENESSRLIARAPKILHFPVITPCTLSSTGSSLIYSDVPEISSRSOLSALQ  
LDYRLLHKSXYIORHIIPEPVITTRPLAKNEDHYSIDAEHSLNLRKIDIT  
SVLSKTRPFRPIHAKLMTHTHLSNTEKSPFACELPFGKMTVSLDAEMIK  
KHCDHMINIEMWINDNVAMTLESILHRSYSLICDRENFITPGRIDOLLDSPLS  
DHRNLIIGKLFVRPHPCWNYRLSLRYGMHRODPNPLYEYOMIILTKIFEHMOIYGYVE  
RRADSSFFFEFLKLDKPKRPPF"

BASE COUNT 2399 a 1827 c 1667 g 2375 t  
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 8268;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcttgatttgaagaaagtcacatgagatggtttataatgctcaaggttcctcatcc 60  
Db 4207 GTGGCTTGATTGAAAAAGGTCATGATGTGTTATTAATGTTCAAGGTCCTCCATACC 4266  
QY 61 aaacatgaaatactgctagagagattgaaacatcgatcctatggactatccatcaac 120  
Db 4267 AAACATTGAATACTTCTCTAGAGAGATTGAACATCGATCTATGGACTATTCATCTAAC 4326  
QY 121 tctcttggacttatttggtagtcttccatcttcttggtagtcttgaatccctgaagtc 180  
Db 4327 TCTCTTGGACTTATTATGTGTACTCTCCATTTCTTCTGTCTTAATTCCTCGTAGTC 4386  
QY 181 tgaagacataaagtccttataataagtaagcaatcacaagagagccttagacaata 240  
Db 4387 TGTAAGCATTAAGATCTTATATTAATAGTACCAATCTAACAGCAGGCTTAGCAACATA 4446  
QY 241 tctgaagacataagtccttataataagtaagcaatcacaagagagccttagacaata 300  
Db 4447 TCTGGAAGGCTCTAATGCTTAGAGGTCATGATGACTTCTTCTTTTAAAGAAAATCGGAGA 4506  
QY 301 agactatcacaagaaagcatccatccatgaatcgcgaacatagaagaaagcacatcat 360  
Db 4507 AGACTATCTCAAGCAAGCAATCCACTCCTCAGATCCGCAAACTGAAAAAGCACATCAT 4566  
QY 361 tgaagagagcctgagagatccttgaagaagcctggaagctgctccccaagctatgaaac 420  
Db 4567 TGGAGCAGGCTGGCGGATCTTCAGAAAGCTTGAGAGTGTCTCCCAAGTATGGAAC 4626  
QY 421 tgcagacccctgcaagcagctactggtttatcggcagctcgaagacatcttggaaaac 480  
Db 4627 TGCAGACCCCTGAGAGAGCTACTGGTTTATCGGAGTCTCAAGACATCTTGGAANAAC 4686  
QY 481 tctgagacactactggtttaaagctttagacatcccatcctgltcatccgcttagaagc 540  
Db 4687 TTTCTGACGACTTACTGTTTAAAGCTTTAGCATCTCCCTATCCTGTCACTCGGCTTAGAGAC 4746  
QY 541 cgcctatagacttgctaaatttgaagaacactaaagtaattgatactatcatcttcat 600  
Db 4747 CGCTTATAGACTTGTAAATTGGAAGACACTAAAGTATGATCATCTACATCTTTCAT 4806  
QY 601 tcaataagctccgaagaataatcgaatgctatctgagcaatatctctaagcttggagac 660  
Db 4807 TCAATAAGCTTCCGAGAAATAATCAATGCTATCTGCGGCAATATTCCTTAGGCTTGAGAC 4866  
QY 661 tgaagaatctgactatatactcgggactctcttagctgccaagaagaagcgagcttggag 720  
Db 4867 TGAAGAATCTGATCTTATATTGGGATCTCTTAGCTGCGCAAGAAAAGCGCATTCGGAG 4926  
QY 721 tgcacagacttgacagatcggaataatacaaaaaagcttcttccgaacacttgaagaa 780  
Db 4927 TGCACAGACTTTCAGATCGAGATACCAACAAAAAGCTTCTTCCGACACTTAGGAA 4986

QY 781 ttgttaacgagtgctgctcctcaagatcaagaagctatctttagctttagggaagct 840  
Db 4987 TTTGCTAACGAGTGGCTGTCCCAAGATCAAGAACCTATTCTTTATGCTTTAGGGAAGCT 5046  
QY 841 taagatgctcaagactcctcaataataaaaaaagatttgaagagccttgatgtagt 900  
Db 5047 TAAGATGCTCAGACTCTACAAATATAAAGCAATGCGAAGCCTGATGTGATGT 5106  
QY 901 caacttaagcaagcttcaagctttaaacttgcttggggaagaagaagcgcctctccgct 960  
Db 5107 CACTTACACAGACTCAAGCTTTATATGCTTTGGGGAAGAAGAGAGAGCGCTCCCGGT 5166  
QY 961 gataaaaaagaagacttggagagagcctcgcagcctctgtagcttaagcatcacc 1020  
Db 5167 GATAAAAAAGCAGACCTTGAGAGAGCGGCTCGAGCCCTGTATGCTTAGCGCATCTAC 5226  
QY 1021 ctctgagaataggattccgatctgcccctcgatatctcctaaataaagaagaagagagc 1080  
Db 5227 CTCTGAGATTAGGATTCGGATTGCCCTCCGATATTCTTAAATAAGAACAGCGAAGC 5286  
QY 1081 caagltgaatgtagctttagctcctttagagttgaagtgtagacacccctaactatgta 1140  
Db 5287 CAAGTTGAATGTAGCTTATAGCTCTTATAGATTAGGGGTGACACCCCTAAACTACTGA 5346  
QY 1141 ataatatcagaaagagcttgcacaaacacatataatgaagcttagcgttagcttctc 1200  
Db 5347 ATAACTTACCGAAAGGCTTGTCACACACATTAATATAGACTTAGGCTTAGAGTTCTC 5406  
QY 1201 taaggggcttacttacaataattgaagcgggtgaacatactagtcctcaagaatcccca 1260  
Db 5407 TAAGGGGCTACTTTACAAATTTGGAACGGGTGAACATCATAGTCCCTCAAGATCCCA 5466  
QY 1261 ggagaaggaaaggttgcctccacaacccgaagctcttgaagaagaatccttaagttct 1320  
Db 5467 GGAGAGGAAAGGTTGCTCTCCACAAACCGAGGCTTGAAGAGAGATCTTACGTTCT 5526  
QY 1321 ctccgcctactcaaaagaagcttaccctccctgattatagaagctttaggagatcgaa 1380  
Db 5527 CTTCGCTACTCTTAAGAGAGCTTACCTCCCTGATTATTAAGCTTTTGGGAGTCAGAA 5586  
QY 1381 aactcaagcttgcacactactgagatctcttctttagtcaacacccatacagaagcctt 1440  
Db 5587 AACTCACGTTGCCACTGCTGGGATTTCTTTTAAAGTACACCCACATCAGAAAGCCTT 5646  
QY 1441 agatcactcttccaaagctcgaaagctcctcggaacactatcatccgcgactatgaga 1500  
Db 5647 AGATCTACTTTTCCAAAGCTGCGAAGCTTCCGGAAGAACTATCATCGCGCCTATGAGA 5706  
QY 1501 tctgctattatataatccaaagaatccctgaaaaaaaagcttctccatgatatgac 1560  
Db 5707 TCTTGCTATTATATATCTCAACAAAGATCTGAAAAAAGGTTCTCTCCATGATTATGC 5766  
QY 1561 aaaaaagctaatcaagaaaccttgcatttgttgaacacggaacacaaagacccatcc 1620  
Db 5767 AAAAAAGCTAATTACAGAAACCTTGTTATTTGTGTGACACGGAACCAAGAACCCCATCC 5826  
QY 1621 caagcatgcccctatctacgttatacagtgcaaccccaagaagccgttagaagctcatgttga 1680  
Db 5827 CAGCATGCCCTATCTAGTATACGTACAGTACACCCCGAAGCCGTCAGAAAGCTATGGA 5886  
QY 1681 taitctagaagaacatagcaacactcgaagcttctccgaagatatccgttatgtatacaact 1740  
Db 5887 TATTCTAGAGACACTACACCTCGAAGTCTCGAAGATATCGTTATTTGATACACT 5946  
QY 1741 gatgaaggaaggaatgcaaaaaattcccaagtccttgagagcttactcotaataatgt 1800  
Db 5947 GATGACGGAAGGAGATCAAAAAATTTCCAGTCTTGAGGCTTACTCAATAAAAAATGT 6006  
QY 1801 ggaagtaaccccaacactacgtctatagaagcttgctctcatcttctagctccctgttc 1860  
Db 6007 GGAAGTAAACCCCAACCTATAGCTTATGAAAGCTGCTTCTTATTTCTGACTTCCCTTGTTC 6066  
QY 1861 ttatgggttctcgaagctgtagctttagatcatcaagaagcgttgaagaa 1907



PSPLQATEKRVSELPFSLFTEKHEVMYIAKLJENGKAKRTIKRCLSIKSFAYHC  
VIOKILLENPAETIHGPRIPKPELSPMYIAQVEVLMATPDLSKHGRLDRIMELPFS  
SGRISEIVANNKDEPDLSTLIRIKGKKERILIPYNSAIQWIOIYLNHPDRLE  
KDPQAFILNFRIRISTRSIDRSFQELRNSGLITPHIRHTIATHMLESGMDLK  
TIOALLGHSSLETTIVTOVSKLQKOTHOEAPHHA"  
BASE COUNT 4710 a 3330 c 3665 g 4743 t  
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 16448;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 gtaggtgatttgaagaagtcacatgagatggttataagttcaagtcctccatcc 60  
DB 13290 GTGGCTTATTTTGAAGAGGCCATGAGATGTTATTAATGATCAAGCTCCCTATCC 13291  
QY 61 aaacatgaactctgctagagagatgaacatcgatcctatgagacattccatcac 120  
DB 13230 AAACATTGAATTAAGTCTAGAGAGTTGAACATGATCTATGAGACTTTCATCTAAC 13171  
QY 121 tctcttggaacttatttctgtagtcttccatctctctctgttgtaaatccctgagtc 180  
DB 13170 TCTCTTGGAAGCTTTATTTGTTAGTCTCCATTTCTCTGTGCTAATTCCTGAGTTC 13111  
QY 181 ttaggtcataagatcccttatataagtaacgaatctacagaagagccttagaaca 240  
DB 13110 TGTAGGTATATAGATCTTTATATATAGTACCAATCTACAGCAGCGCTTAGCAACATA 13051  
QY 241 tctggaagctctagatgctcagtgatcagatctctgcttcttgaagaataatcgagga 300  
DB 13050 TCTGGAAGCTCTAGATGCTCTAGCGATGATGATGATCTCTCTGTTTAAAGAAAATCGAGA 12991  
QY 301 agactatctcaagaagaacatccatccctcagatccgcaaatagaaaaagcacatcat 360  
DB 12990 AGACTATCTCAAGCAAGAACATCCACTCCTCAGATCCGCAAACTGAAAGACACCATCAT 12931  
QY 361 tggagcagagcctgagcagatcttcaagaagccttgagcgtgctcccaagctatggaac 420  
DB 12930 TGGAGCAGGCTGCGGGGATCTGAGAGCCTTGAGACGCTGCTCCCAAGCTATGGAAC 12871  
QY 421 tgcgaacccccctgcaagcagctacgtgttatacggcagctcagagacatcttggaac 480  
DB 12870 TGCAGACCCCTGAGAGCAGCTACTGTTTATCGGCACTCTCAGAGACTCTTGGAAGAAC 12811  
QY 481 tctctgagacttactgtttaaagctttagcatctccatctcgtlcatcgccttagaagc 540  
DB 12810 TTTCTGAGGACTTACTGTTTAAAGCTTAAAGCTATCCCTATCTGTCATCCGCTTAGAAGC 12751  
QY 541 cgctatagacttcttaatttgaagaacacaaagtcatttgaatcatctacattcttcat 600  
DB 12750 CGCCTATAGACTTCTAATTTGAAGAACACTTAAAGTCAATGATATCTACATTTTTCAT 12691  
QY 601 tcatatgctcccgaaagaatccaatgcctatctcgcgcaatcttctcagccttgagac 660  
DB 12690 TCATTAAGCTTCCGAGAAATCCAAATGCTATCTGCGGCAATATCTTACCTTGAGAC 12631  
QY 661 tgaagaatctgagtgcttatatctgggagctcttagctgccaagaagaacgagatcgag 720  
DB 12630 TGAAGAATCTGATCTTATATTTCCGGGATCTCTTAGCTGCGCAAGAAAGCGGATTCGAG 12571  
QY 721 tgcgaagccttgagatcgagagatacacaagaagaagccttcttcgcaacttagaa 780  
DB 12570 TGCCACAGCTTTGCAAGATCGGAGATACCAACAAAGCGCTTCTTCGACACTTAGAA 12511  
QY 781 ttctgtaacgagtgctctcctcaagaatcaagaagctatcttcatgtcttagaagc 840  
DB 12510 TTTGCTAAGCAGTGTCTCTCTCTCAAGATCAAGAGCTATTTCTTTATGCTTAGGAGACT 12451  
QY 841 taagatggtcagaagctactacaataaaaaagcaattgcagaagcctgagtgtga 900  
DB 12450 TAAGGATGCTCAGAGCTACTACATATATAAAGCAATTCAGAGCGCTGATGTGATGT 12391

QY 901 caatttagcagagctcaagctttaaattgcttggggaagaagaagcgtctccgt 960  
DB 12390 CACTTACAGAGCTCAAGCTTTTAAATTTGCTTTGGGGAAGAGAGAGCGCTTCCGT 12331  
QY 961 gataaaaaagcaagcacttgagagcggcctcagcctgtatgcttagcagcatcac 1020  
DB 12330 GATAAAAAAGCAAGCCTTGAAGAGCGGCTCGAGCCTGTATGCTTACGGCATCTAC 12271  
QY 1021 ctctgagataggaattccgattgcccctgcgatatctcctaaanaacaaagcgaagc 1080  
DB 12270 CTCTGAGATAGGAGATTCGATTTGCGATTTGCCGATTTCTTAAANAACATAAGAACGGAAGC 12211  
QY 1081 caagttgaatgagcttagatctctcttagagttgggtgagacaccttaactactga 1140  
DB 12210 CAAGTTGAATGATGCTTTAGCTCTCTTATGAGCTTAGGGGTGACACCCCTTAATCTGGA 12151  
QY 1141 atacatccgaagagctgtgccaccccatataatagactcagcctgagttctc 1200  
DB 12150 ATACATTCGGAAGAGCTTTGTCACCAACCATTTATATAGACTCTAGCCTTAGATTCTC 12091  
QY 1201 taaggcgctacttacaanaattggaagcgggtgaacatcagtlccctcaagatcccca 1260  
DB 12090 TAAAGGCGGTACTTTACAAATTTGGAACGGGTGAACATCATATAGTCCCTCAAGATCCCA 12031  
QY 1261 ggaagaggaaggttgctctcccaaacccgaggtcttgaagaagaatcccttagttct 1320  
DB 12030 GGAGAGGGAAGGTTGCTCTCCACAAACCCGAGGTCTTGAAGAGCATCTTACGTTTCT 11971  
QY 1321 ctccgcctacttaagaagcttaccctccctgcttatataagcttgggagatcaaga 1380  
DB 11970 CTTCGCGCTACTTAAGAGAGCTTACCTCCCTGTATTTATAGCTTTTGGGAGTCAAGAA 11911  
QY 1381 aactcagcttgccactactgagattcttctttaaagtcacacatcaacaggaagcct 1440  
DB 11910 AACTCAGCTTCCCACTACTGCGATTTCTTTTAAAGTACACACCTACATCAGGAAGCCTT 11851  
QY 1441 agatctacttctcgaagctggaagctcctggaagaactatcatcgcgctatgaga 1500  
DB 11850 AGATCTACTTTTCCAAAGCTGGGAAGCTTCCCTGGGAAGCTTATCCCGCTATGCGA 11791  
QY 1501 tcttgctattataatctcaccaagaagatccctgtaaaaaaacgttctctcctatgata 1560  
DB 11790 TCTTGCTATTTATATCTCACCAAGATCTCGAAAAAAGCGTTCTCTCCATGATATGC 11731  
QY 1561 aaaaaagctaaattcaaggaacactgttatttgytgaacggaagaacaaagaccatcc 1620  
DB 11730 AAAAAGCTAATTCAGGAACCTTGTTATTTGTGACACAGGAAAAACCAAGACCCATTC 11671  
QY 1621 cagcatgctctatctacgttatacagtgatcccccagaagaacgctagagctcatgtgga 1680  
DB 11670 CAGCATGCCCTATCTAGCTTATCAAGGTACACCCCGAAGACCGTACGAAGCTCATGTTGA 11611  
QY 1681 tatctagaagacatagcacaactcgaagctcttcgaagatatcgttatatgataact 1740  
DB 11610 TATCTAGAGACATAGCACTGCAAGTCTTCGGAAGATATCCGTTATTTGATACACT 11551  
QY 1741 gatgacggaagagatgcaaaaaattccagatccttgcaaggttactcatlaaanaatt 1800  
DB 11550 GATGACGGAAGAGATGCCAAAAATTTCCAGTCTTGCAAGGCTTACTCATAAAAATTGT 11491  
QY 1801 ggaagtaaccccaactcagctgttataagaaogttgcttcttatctcgtctccctgttc 1860  
DB 11490 GGAGTAAACCCCAACCTTACGTTATGAAAGCGTTGCTTATTTCTAGCTTCTTGTTC 11431  
QY 1861 ttatgggtctcctcagctgagcttctgcatcaagaagcgtgtgaa 1907  
DB 11430 TTAAGGTTCTCTCAGCTGATGCTTTGACATCAAGAGGCTGTGAAA 11384

RESULT 4  
AX349527 1707 bp DNA linear PAT 06-FEB-2002  
LOCUS AX349527  
DEFINITION Sequence 50 from Patent WO0202606.

ACCESSION AX349527  
VERSION AX349527.1 GI:18615370  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae.  
ORGANISM Chlamydomophila pneumoniae.  
REFERENCE  
1 (sites)  
Ratl, G. and Grandi, G.  
TITLE Immunisation against Chlamydia pneumoniae  
JOURNAL Patent: WO 0202606-A 50 10-JAN-2002;  
Chiron S.P.A. (IT)  
FEATURES  
source  
1..1707  
/organism="Chlamydomophila pneumoniae"  
/db\_xref="taxon:83558"  
BASE COUNT 493 a 415 c 340 g 459 t  
ORIGIN  
Query Match 89.5%; Score 1707; DB 6; Length 1707;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 atggagactatccaactcaactcctcttggaactttatgtgtagcttccattctct  
1 ATGGAGACTATTCACATCACTCTTGGACTTTATGTGAGCTTCCCATTTCTCTT  
Db 161 gtgtgtaattccctgagtctgtgagtcataagaatcccttatataagaatgaatctaca 220  
61 GTTGTCAATATTCCTGAGCTGTAGTCATAGATCCTTATATATAGTACGAATTCACA 120  
QY 221 cagaagccttgaagaacatacttggaagctctagaatcctaagtgatcactcttc 280  
121 CAGAGGCTCTTAGAACAATCTGAGAGCTCTAGATGCCCTAGTCAAGATTTCTTC 180  
QY 281 gttttaagaanaatcggagaagactatcacaagaagaatccaactccatgaatcgcaa 340  
181 GTTTTAAGAAAATCGAGAGAAGACTATCTCAAGCAAAAGCATCCACTCAGATCCGCAA 240  
QY 341 actaagaanaacaccatcatttggaagagcgtgagcagactcttaagaagccttgaagtg 400  
241 ACTAGAAAAGACCACTATTTGAGAGCGCTGGCGGAGTCTTGAAGCCTTGGACGTG 300  
QY 401 cctcccaagcctatggaactgcagacccctcagcagcactgactggtttatcgagc 460  
301 CTCGCCAAGCTATGGAACATGCGAGACCCCTGACAGCAGCTACTGTTTATCGCGAGTC 360  
QY 461 tcaagacactcttggaanaactctgcagactactglttaagccttgaacatccctat 520  
361 TCAGGACATCTTGGGAAAACCTTGCAGCAGCTTACTGTTTAAAGCTTTAGCATCTCCCTAT 420  
QY 521 ccgtgtatccgcttagaagccgcctatagactgtgtaatttgaagaacacataagctatc 580  
421 CCTGTCAATCCGCTTAGAAGCCGCTATAGACTTCTAATTTGAAGACACTAAAGCTATTT 480  
QY 581 gatcatctacatcttcatcatcaagcttccgaagaanaatccaatgctatctgcgca 640  
481 GATCATCTACATCTTTCATTCATTAAGCTTCCCGAAGAAAATCCATGCTTATCTGCGGCA 540  
QY 641 atattcctcagcttggaagactggaatctgaatctatatatctcggatctctgaatgccc 700  
541 ATATTCTACGCTTGGAGACTGGAAGAAATCTGATCTTATATTCGGGATCTCTTAGCTGCC 600  
QY 701 aagaagaagcagatctggagtgccacagccttgcagatcggagaataaccacaagaagc 760  
601 AAGAAAAGCGGATTCGAGAGGCCACAGCTTTGACAGATCGGAGAAATACCAACAAAAGCGC 660  
QY 761 ttctctcgaacacttaagaattgtctaagagtgctcctcctaagaatcaagaagcatt 820  
661 TTTCTTCCGACACTTAGGAATTTGCTAAGAGAGTGCGTCTCTCAAGATCAAGAAAGCTATT 720  
QY 821 cttaatgcttaagggaagcttaagaatggttcaagaactactacaataataaagaagaatg 880  
|||||

Db 721 CTTTATGCTTTAGGGAAGCTTAAGGATGTCAGAGCTACTACATATATMAAAGCAATTG 780  
QY 881 cagaagcctgatgtgagatgctcactttagcagcagcctcaagccttaattgtcttgaggaaa 940  
|||||  
Db 781 CAGAAAGCTATGTGATGTGACACTTTTAGCGAGCAGCTCAAGCTTTAATGTCTTTGGGAAA 840  
QY 941 gaaaggaagcctctcccgatataaaaaaagaagcacttgaaggagcgccctcgagccctg 1000  
841 GAAGAGAGACCTCTTCCCGATTAATAAAGCAAGCACTTGAAGAGCGCGCTCGAGCCCTG 900  
Db 1001 tatgcttaagcgaatctaccctctgagatagggaattcgaatgcctgcgcatctccta 1060  
901 TATGCCCTTAGCGCATTCACCTCTGAGATAGGAGTTCGATTCGCCGATATTCCTA 960  
QY 1061 aaactaagaagaagcgaagcgaagtgaatgtagcttgaactcctccttagagttagggtgt 1120  
961 AAAACTAAGAACAGCGAAGCAAGTTAAATGTACTTTAGCTCTCTTAGAGTTTAGGCTGT 1020  
QY 1121 gacacccctaaactactcgaatacatcaccgaagagcttgccaaaccacatataatgag 1180  
1021 GACACCCCTTAAGCTAGTGAATACATTACGAAAGGCTTGTCCACACATTTAATGAG 1080  
QY 1181 acttgaagccttgaatctctctaaggggcgaacttacaanaatggaagcggtggaacatc 1240  
1081 ACTTGAAGCTTGAAGTTCTCTTAAGGGGCGTACTTACAAAATTGGAAAGCGGTGAACATC 1140  
QY 1241 atagtccttaagaatcccgagaggaaggaaggttgcttcgaacacccgaaggtctgaa 1300  
1141 ATAGTCCCTCAAGATCCCGAGAGGGAAGGTTGCTCTCCACACCCGAGGCTTTGAA 1200  
QY 1301 gaagcagatccttaagcttctctcctccgactaactaaagaagcttaccctccctgattat 1360  
1201 GAGAGATTCCTTAGCTTTCCTTCCGCTACCTAAGAAAGCTTACCTCCCTGATTTAT 1260  
QY 1361 aagcctttagcgagtcgaanaactcagactgagcactcgtgagatcttctttaaagtcac 1420  
1261 AAGCTTTTGGGAGTTCGAAAAACATCAAGCTTGCACATCGATGATTTCTTTTAAAGTCAC 1320  
QY 1421 accctacatagaagaagccttagatcacttctcgaagctcgaagcttccctgagaacct 1480  
1321 ACCCTACATCAGAAAGCCTTAGATCTTCTTTCCAAAGCTGGAAGCTTCCGGAAGACT 1380  
QY 1481 atcatccgcgactatgagatctgtcatlataatcaccacaagaatcctgaaaaaaa 1540  
1381 ATCATCCGCGGCTATGAGATCTTGTATTATATCATCCACCAAGATTCGAAAATAAAA 1440  
QY 1541 cgttctctcaatgatatgcaaaaaagctaattcagaagaacctgttatttgtgagacg 1600  
1441 CGTTCCTCCATGATTAATGCAAAAAGCTTAATTCAGGAACCTTGTTATTTGTGGACAG 1500  
QY 1601 gaaaccaaagaagcccatcccaagatgcccatactagctatcagatcaagtcaccccaagaagc 1660  
1501 GAAAACCAAAAGCCCATCCAGCATGCTTATCTTACGTTATCAGGTACCCCGAAGAGC 1560  
QY 1661 cgtacgaagctcagctttagatattcttagagacactaagccactcgaagcttctcgaagat 1720  
1661 CGTAGAAGCTCATGTGGATATTCFAGAGACACTAGACACTCGAAGTCTTCGAGAGAT 1620  
Db 1721 atccgcttatatgatacaactgagatgaggaagagatgcgaanaatttcccaagctctgca 1780  
1621 ATCCGTTTATGATTAACAATGATACGGAAGAGATGCAAAAATAATTTCCAGTCTTGCA 1680  
QY 1781 ggcttactcaataaattgtgagtaa 1807  
1681 GGCTTACTCATTAATAATTGTGAGTAA 1707  
Db 1807  
RESULT 5  
AE001308 11627 bp DNA linear BCT 30-Oct-2000  
LOCUS Chlamydia trachomatis section 35 of 87 of the complete genome.  
DEFINITION AE001308 AE001273  
ACCESSION AE001308.1 GI:3328766  
VERSION



Db	4512	TTCTGAGACTTCAGCAACTTATCAACAACAGACATTTTCCTCCGATTATACCGGTAATATGACAGA	4571
Qy	301	agacatctccaagaacacatccatccctccgatcccgacaacttagaaaaagccatcat	360
Db	4572	ATCGATTATTACACAACACCTTCTCTCGAGSACACGTACATACGTAAAGGCAATTAT	4631
Qy	361	tggagcaagcctcgcgggactctcgaagaccttgaagctgcctcccaagctatgaaa	420
Db	4632	TGGACACAGGGCTATCTGGTTTCATCGAAGCGTTTAAGATTACTGCTCGAGGGCATATGAAAC	4691
Qy	421	tgcgagcccccgcagcagcctactcgttttatccgcagctccagacatctcggaaaac	480
Db	4692	GCAAAATCTCTATGAGCACTACTATCTTTTAATCTCTCAACCAACCAATTAAGCAAAAC	4751
Qy	481	ttctcagacttactcgtttaaagcttagacatccctacccctcgtccacccgcttgaagc	540
Db	4752	TTCTGACAAACTTATTATTCAGAGGATTTAACAGCTTCTCATCTCGTCACTCCGTTAGAAC	4811
Qy	541	cgcctatagacttgcttaatttgaagaacactaaagctatgatactacatcttcaat	600
Db	4812	TGCTTATGCTCTGCCCTGATGCAAAAATAAGCAAGTAGTGATGATACCTTATCTTTAT	4871
Qy	601	tcatagactcccgagaagaatccaatctgcatactcgcgcaatctccagcttggagac	660
Db	4872	CTACAGTTTACCGAGAAATTCAAACCTACGGCACTATTTCTTACCACTCGAAAC	4931
Qy	661	tgaagaatctgattgcttatatccgatctctcctagctgcgaagaaagcgacttcggag	720
Db	4932	AGAAACAGCTGATGCTTATATTCAATCATTTGCTCTCTTCTCCCAATTAACCTACAGAAA	4991
Qy	721	tgcacagccttcgcagatcgcgagaataccacaacaaacgccttcttcgcacactaggaa	780
Db	4992	CTATGTGCTATTTAAATTGAGACTACAAACAAAAGATTTCTTCAACACTCGCGTC	5051
Qy	781	tttgctaacgagtgtgctctccctccaagatccaagaactatcttcttgctttaggaaagt	840
Db	5052	TTTACTTTCACAACTGGCTCTCTTGTGATCAACAAAGCGCTTTGTATGCGATGAGCAAACT	5111
Qy	841	taagatagtccaagactactacaataaaaaagaacattgcagaagccctgatgtgagt	900
Db	5112	GGAGACTCTGGTACTATCTTAGAATTTAAAGCTTTAAGCTTAAGCTTAAGTCCAATCTGAAAT	5171
Qy	901	caactttagcagagctcctaagcttlaattgcttgggaaagaagagacgcctctccgct	960
Db	5172	AGTACTCGCTGAGCTGACGACTATTATTCTTACGAAAGAAAGAAAGAGCTCTACCGAT	5231
Qy	961	gataaaaaagcaagcactcttagagagcggcctccgagccctcgtatcgccttagcgacatcac	1020
Db	5232	CCTAACCAACCTTTCGCAACAAAACCTTCTTCGACCCCTGTATPACCGCACCTTCTCTTC	5291
Qy	1021	ctcttagatagagatccgatctgcctcgcgcgtatctcccaaaactaagaagaagcagc	1080
Db	5292	GCAACAGAAAGGTGAAAGCTTCTTCTTCACATCTTTTATTAACGCAACACAGAAAGAAAT	5351
Qy	1081	caagtttgaatgactcttagctctctctliagagttagagtgtagcaccacctaaactactgga	1140
Db	5352	TAGACTGAAATACGTGTTTAGACACTGTTTCATCAAGAGGTGTACAGATTCCTCAAGTCTCTCA	5411
Qy	1141	atacatccogaagagctctgccaacacatataatgaagctctagccttagagttctc	1200
Db	5412	CTATCTAACAGAAATCTTAGAAGATTAAGATTTCTCATGCAATATTTTACTATACACAC	5471
Qy	1201	taagaggcgctactttaacaaaattggaagcggtgtgaacatcatagctccctcaagatcccca	1260
Db	5472	GACAGGAAAGCTATACAGTTCTTGAAAGAAATGACACCACTTTCCTTCATGAGCCAAAGA	5531
Qy	1261	ggaaggggaaagtgcttcctccacaacccggaagctcttgaagcgagatcccttagcttct	1320
Db	5532	AGACAAATGAGAACGTTGGCTATGTATTCGGGTACCGGAAGATACCATCTCTCAGCGTT	5591
Qy	1321	ctccgcctaccataagaagcttacctccctcgtatattataagcctttggcgagctcagaa	1380
Db	5592	ACTAAATTTACCAATGAGCGCTATCTTCTTACTACGTAGAGGCCATCTCGGCTTCACAAA	5651







5284. .6870  
/gene="ABC transporter, ATP-binding protein; identified by match to PAM protein family HMM PF00005"  
/codon\_start=1  
/transl\_table=1  
/product="ABC transporter, ATP-binding protein"  
/protein\_id="AAF73579.1"  
/db\_xref="GI:8163271"  
/translation="MSIVLDKIGKTLGRVLEFDVSVFENGNGNYGLTGRGACKSTL  
LKITGVPEPSGCTISLPKKGIGLRQNDISBVDVLDVYMGNSRLMEAMQRDADY  
AEEFDVAGMEIGETIELDIGEEDVRASEAEELLIGIPEEFSOKMATTPLDLP  
RVLLQALFGEHPEALDEEPNHLDSIMNIGNLFMDYNGTVIVSHDRFLITFT  
HADIDYPTIIIPGNYDAMVEMKTAASEQEKADIKSEKELIADREFVAFGSGSA  
SOVSORLEIKELIPOEKLKSNIOBPYRFPVSQASGVYESTIGSKTYADAPLEN  
PSEIIEGDKIVGIVGNGGICKTLMKILAGVEPSSAQIKLGNVAFSPFQNHADY  
LRDCEETLFEMLRRRKTCITTDQETRSVLGKMGDGAERQVALSGGEFARLLMG  
MLLENHNLVLEDEANNHLDLDESVALAMAINDPGTAIFVSHDRTLIECATKLLIFD  
KGVISFPGPMSVYLAASKK"  
6903. .7493  
/gene="TC0628"  
6903. .7493  
/gene="TC0628"  
/note="similar to GB:L08793 SP:Q02169 GB:M96343 PTD:142853  
PTD:143161; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="maf protein"  
/protein\_id="AAF39457.1"  
/db\_xref="GI:7190666"  
/translation="MGTVLVGGSSSKIRKAVLEAFRIPEICVSDPERSITYSGDPE  
ETRLRANCKANAASVSGESDLSIITADTVVYVGEVGNPKRESEHAEVEMLRITSGTS  
HSVITSLVMNEKVASASEPTQVSPFDIPQHLKTYQAFSLKRCGGYCYVDGSGSL  
IKIQEGCVYINQGIPLTLMQLMERNISLMDYLA"  
7473. .9175  
/gene="TC0629"  
7473. .9175  
/gene="TC0629"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAF39458.1"  
/db\_xref="GI:7190667"  
/translation="MGLSRLIFELSLSPASGCDPPPSYSQKLEFLCQSTPOALES  
YLEASTTQQNHFSLIRLAKSYLQGLSESDAVVRSAIIAGLSGSSLETLDLSES  
LEFDQVYHQLILNAGNQLGKTSDRLEFKGLPAPRILEAAVRLACMKNKNSVDY  
LYSFHOLPEBRIOLANILFQLETEADAVHRLSSPNSLTNNYAVYVIGEVOQR  
FLPRTSLTSLTAAPLDOEGSLYAIKLEDASGYIKALSRSYVAAALQOTILEFL  
KREDALSLITTFPOOGLPRAIVYSRFLSESGEELLIPCKAKREITANMAIYAL  
HGSVNNHLYSLITFELENKILIHRIPLPHSGIKATOPKKECTPLPLSPEKARALA  
MYRAEDPELKLTPNNAYLPYLERLISQKTPLAKAATAFLSVTAHQALSLVK  
AALVGGDTIRAVANLALYMTQDPEKKALLYQAEQLIDGTTLTDEENPLSPSS  
YLRVQSEPTRSQMLTLTLELVSSKDEDEDIRVPLSLMKRTHYKNIPILSLMLRIVE  
"

gene  
CDS  
9182. .11284  
/gene="TC0630"  
9182. .11284  
/gene="TC0630"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAF39459.1"  
/db\_xref="GI:7190668"  
/translation="MKRFFLFCVILIAHNSPEGLSHOQAIQKISYLSHFHGITGI  
MDVEGVGHIHDDLRLQANKKAYVENRDCGKIIQAVHNVNNGKILIDYEDSD  
TDSCLIVGRCOSLIPWFIIGSSTNTISPSIIIGKYISTEGSPQKHCLSGDLVSS  
ESVLSMGPTRLSICNTPLPIPTQFSINPMETPKRPTFRFGSGSGFAGSYLGISTPS  
KKHSTLTFLDSFRHGIIGLGNIRFSSOENPGNAVIRKTYAHRIALIDSSGADRAR  
LHGDAFSKTKTRLISGEPHLSDSWEYVADIFPNNVSLKNTGPEIHLIGWDNDPEFV

		TSYKVSFNQVKQLPOVILHHPNVINIGSRITLENRIEAGFLDHFSDNLTGNSFSSWRLSSHKRIFERSLALPDIQTLPILSGTAIFYSSMLQONAHQAOSLSLDYRPVLQKEYRHARHIVEPCFSFKTTPDLSPSGETAFYSKIDMASHNLLOIQIEFSEVYLKHSPTPIASHLKIWTCTJEDPEVATKTETPKPAKWHSIPLKDANLONTSTIDEAMTKMRBDHNIPIMWIINDNIGTLFELEPHRSKVFICACADKNYLDLDSPLKTIASPLSGRRLLTGKEFRPHPHMNYNLNIRKGWRHPSPLYELOMILGHKIFEHMOISVEYEREADKRRCFFYLKDKRKHKHCHPRG"
gene	CDS	complement(11348..11656) /gene="TC0631" complement(11348..11656) /gene="TC0631" /note="conserved hypothetical protein; identified by GIImmer2; putative" /transl_start=1 /product="conserved hypothetical protein" /protein_id="AAF39460.1"
Query Match	Best Local Similarity	26.8%; Score 511.8; DB 1; Length 11765; 55.6%; Pred. No. 1.4e-130;
Matches 1006:	Conservative	0; Mismatches 797; Indels 6; Gaps 1;
Oy	1	gtgccttgaatttgaaaaagtcocatggatgltttataatglttaaggctcccatcc 60
Db	7375	GTGGTTATCATCAACAATAAGTAGAGGTTGGTATTAAATAACAGGAGATTGCCATTAA 7434
Oy	61	aaaccltgaactactgttagagaggttgaaaatcgtacctctctggacatccatcac 120
Db	7435	AGACCCTAACCAGTGGCTCATGGAGTTAACATACGTTATAGGGATTATTCGGCTATAT 7494
Oy	121	tctcttggacttttatgtgtgtagcttcaccatctctctctgttgtcaaatccccgagtc 180
Db	7495	TTTATTG-----CTTACTTCTTTHACCGCTCCACACAGCTCCGACTTCCCCCAG 7548
Oy	181	tgtagtcatagaatcctctatatagaacgcaatctaacagcaggcgcttagacaata 240
Db	7549	TGTTTTCCAGAGAATATTATTTCTGTGCACAAAATAATCTCCTCAAGCTCGAGTCCCTA 7608
Oy	241	tctggaagctctagaatgctcagtgatcatgacctctcgtttaagaaaaatcggaga 300
Db	7609	TCTTGAGGCATCTACACCTTATCAACAACATTAACCTTTCCTATATTTGCCCTTAATAGCTAA 7668
Oy	301	agaactatccaagaacgacalccactccctcagatccgcaaatagaaaagacacatcat 360
Db	7669	GTCATACTTACACAAATGCTCTTTCTTGAGAGTGTCTTACGTACGCCAANAAGCGCATATAT 7728
Oy	361	tggagcaggcgctgcggagatcctcagaagccttggacgtgctctcccagatbatgaaac 420
Db	7729	TGGAACGGGGGCTTTCTGTGCTCATCTGAGACCTTAATGATCTGTGATCATATGAAAC 7788
Oy	421	tgcgaagccctcgagcgctcactggttttatcgcgagctccagaaactcttggaaaac 480
Db	7789	ACAGGATTTTATGAGCAGCTACTATTTTTAATCTCCAGGCATCAATTAGGCAAAAC 7848
Oy	481	ttctgaagacttacgttttaaagcctttagacatcccatccatcctgycalcgcgttgaagc 540
Db	7849	TTCGATGTGCTTTTATTCAAAGATTTAACAGCACCTCATCTTATATTCCGTTGGAGAC 7908
Oy	541	cgcctatagacttgciaatttgaagaacataaagtcattgatcatccatcttccat 600
Db	7909	TGCTTACCGTCTGGCGCTGTATGAAAAACAGTAAGTAGTACATCTCTATTCTTTTAT 7968
Oy	601	tcataagcttcgcgaagaatccaatgctctatctgcggcaatatctccagcttggagac 660
Db	7969	CCACCGCTTCACAGAGAATTCAAAATTTAGCAGCAACGATTTTTTTTGACGCTGGAAC 8028
Oy	661	tgaagacatgatccttatctcggaatcctctttagcctccaagaanaagcgcattcgag 720
Db	8029	GGAGAAGACCAATGCTTATGTTCATAGACCTCTGTCTTCTCCTAATATATACAGAA 8088
Oy	721	tgccacagcttltcgagalccgagaaataccaaaacgccttctctccgacattagaa 780
Db	8089	CTATATGGCTTATCTAATTTGAGATATCAACAGAGGAGATTTTCTTCCAAGCTCGGTC 8148





FEATURES	source	Location/Qualifiers
		1..128866
		/organism="Oryza sativa"
		/clivair="Nipponbare"
		/db_xref="taxon:4530"
		/chromosome="2"
		/clone="OJ1212_E12"
BASE COUNT	34028 a 30313 c 29226 g 35194 t	105 others
ORIGIN		
Query Match	2.1%; Score 40.6; DB 2; Length 128866;	
Best Local Similarity	52.7%; Pred. No. 5;	
Matches	88; Conservative 0; Mismatches 79; Indels 0; Gaps 0;	
Oy	1332 ccaactacgcgattcctttttaagtcacacctacacacgaagaagccttagatcattt 1451	
Db	5470 CAACCTATGTTATGACATATTTAAGTCACAGCTTCCCTCAGCATGTCGATATCTTTGTG 5411	
Oy	1452 tccaagcgcgaagcttcctcgcgaacctatcaccgcgcctatgcagatcttctattt 1511	
Db	5410 CTGCTGCTTGAGTCGTGATGAACGCACATCTAGAGATCAACAGTCAATCTTGATGAT 5351	
Oy	1512 ataatccaccaagatcctgaaaaaagcttcctccatgatatt 1558	
Db	5350 TTCTAATTAGCAATTATATTGACATTAAGAATTTCTCTATGTTATAT 5304	
RESULT 11	AL445687	168108 bp DNA linear PRI 24-NOV-2000
AL445687/c		
LOCUS	Human DNA sequence from clone RP11-567B20 on chromosome 1, complete sequence.	
ACCESSION	AL445687	
VERSION	AL445687.5 GI:11322702	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE	1 (bases 1 to 168108)	
JOURNAL	Thomas,D.	
COMMENT	Direct Submission	
	Submitted (24-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,	
	CB10 ISA, UK. E-mail enquiries: humquerry@sanger.ac.uk	
	requests: clonerequest@sanger.ac.uk	
	On Nov 23, 2000 this sequence version replaced gi:1113977.	
	During sequence assembly data is compared from overlapping clones.	
	Where differences are found these are annotated as variations	
	together with a note of the overlapping clone name. Note that the	
	variation annotation may not be found in the sequence submission	
	corresponding to the overlapping clone, as we submit sequences with	
	only a small overlap as described above.	
	This sequence has been finished according to sequence map criteria	
	as follows. An attempt is made to resolve all sequencing problems,	
	such as compressions and repeats, but not necessarily within known	
	annotated human repeat sequence elements (e.g. Alu). Where the	
	sequence is ambiguous, there is an annotation using the 'unsure'	
	feature key.	
	The following abbreviations are used to associate primary accession	
	numbers given in the feature table with their source databases:	
	Emn, EMBL, Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP. Information	
	on the WORMPEP database can be found at	
	http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence	
	was generated from part of bacterial clone contigs of human	
	chromosome 1, constructed by the Sanger Centre chromosome 1 Mapping	
	Group. Further information can be found at	
	http://www.sanger.ac.uk/HGP/Chr1	
	RP11-567B20 is from the library RPCR-11.2 constructed at the	
	Roswell Park Cancer Institute by the group of Pieter de Jong. For	
	further details see http://bacpac.med.buffalo.edu/	
	VECTOR: pBAC3.6	
	This sequence is the entire insert of clone RP11-567B20.	

FEATURES	Location/Qualifiers
source	1. 168108
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP11-567B20"
	/clone_lib="RPC1-11.2"
	742..771
repeat_region	/note="15 copies 2 mer ca 100% conserved"
repeat_region	1032..1260
misc_feature	/note="MIR repeat: matches 20. .259 of consensus"
misc_feature	/note="match: STS: Em:G62829"
repeat_region	1546..2029
repeat_region	/note="match: GSS: Em:AQ454440"
repeat_region	2503..2702
repeat_region	/note="MIR repeat: matches 34. .238 of consensus"
repeat_region	3036..3331
misc_feature	/note="AluXg repeat: matches 3. .229 of consensus"
misc_feature	complement(943)..5462)
misc_feature	/note="match: GSS: Em:AQ706543"
misc_feature	complement(501)..5474)
misc_feature	/note="match: GSS: Em:AQ401497"
repeat_region	5481..6203
repeat_region	/note="match: GSS: Em:AQ377544"
repeat_region	6847..6981
repeat_region	/note="L2 repeat: matches 2559. .2710 of consensus"
repeat_region	7538..7661
repeat_region	/note="L2 repeat: matches 2121. .2243 of consensus"
repeat_region	7813..8115
misc_feature	/note="AluXs repeat: matches 1. .303 of consensus"
repeat_region	8126..8494
repeat_region	/note="match: STS: Em:G07568"
repeat_region	8341..8414
repeat_region	/note="RHEB repeat: matches 1. .74 of consensus"
repeat_region	8927..9222
repeat_region	/note="AluB repeat: matches 1. .301 of consensus"
repeat_region	9302..9513
repeat_region	/note="MIR repeat: matches 20. .250 of consensus"
repeat_region	9562..9760
repeat_region	/note="MIR repeat: matches 48. .249 of consensus"
repeat_region	9849..9961
misc_feature	/note="L2 repeat: matches 1149. .1258 of consensus"
repeat_region	10016..10365
repeat_region	/note="match: GSS: Em:AQ276162"
repeat_region	10035..10160
repeat_region	/note="MTL1 repeat: matches 19. .185 of consensus"
repeat_region	10322..10379
repeat_region	/note="29 copies 2 mer ca 75% conserved"
repeat_region	10500..10646
repeat_region	/note="L2 repeat: matches 1373. .1544 of consensus"
repeat_region	10731..10844
repeat_region	/note="MER58 repeat: matches 1. .117 of consensus"
repeat_region	10860..11134
repeat_region	/note="L2 repeat: matches 1556. .1828 of consensus"
repeat_region	11162..11490
repeat_region	/note="L2 repeat: matches 1879. .2234 of consensus"
repeat_region	11564..11790
repeat_region	/note="MSTA repeat: matches 1. .207 of consensus"
repeat_region	11791..12100
repeat_region	/note="AluS repeat: matches 1. .312 of consensus"
repeat_region	12101..12264
repeat_region	/note="MSTA repeat: matches 207. .426 of consensus"
repeat_region	12400..12519
repeat_region	/note="L2 repeat: matches 2626. .2750 of consensus"
repeat_region	12786..13164
misc_feature	/note="MSTA repeat: matches 4. .426 of consensus"
misc_feature	complement(1355)..13937)
repeat_region	/note="match: GSS: Em:AQ116449"
repeat_region	14332..14379
repeat_region	/note="MIR repeat: matches 208. .257 of consensus"
misc_feature	complement(14897..15458)
misc_feature	/note="match: GSS: Em:B48614"

```
misc_feature      15471..16013
                   /note="match: GSS: Em:AQ378181"
misc_feature      15475..15980
                   /note="match: GSS: Em:AQ264491"
repeat_region     15558..15643
                   /note="43 copies 2 mer aa 64% conserved"
repeat_region     17157..17216
                   /note="12 repeat: matches 1734..1793 of consensus"
repeat_region     17221..17278
                   /note="29 copies 2 mer ga 69% conserved"
repeat_region     17442..17658
                   /note="MER20 repeat: matches 3..218 of consensus"
repeat_region     17749..18098
                   /note="12 repeat: matches 2353..2690 of consensus"
repeat_region     19071..19210
                   /note="MIR repeat: matches 78..230 of consensus"
repeat_region     19294..19608
                   /note="AluSg repeat: matches 1..313 of consensus"
repeat_region     19912..20041
                   /note="12 repeat: matches 1627..1761 of consensus"
repeat_region     20452..20492
                   /note="MADP1 repeat: matches 1..42 of consensus"
repeat_region     20493..20789
                   /note="12 repeat: matches 1755..2089 of consensus"
repeat_region     21189..21943
                   /note="L1PA15 repeat: matches 4705..5420 of consensus"
repeat_region     21944..22243
                   /note="AluY repeat: matches 1..301 of consensus"
repeat_region     22244..22959
                   /note="L1PA15 repeat: matches 5420..6155 of consensus"
repeat_region     23511..23788
                   /note="AluYb repeat: matches 11..288 of consensus"
repeat_region     23947..24243
                   /note="AluSp repeat: matches 1..298 of consensus"
repeat_region     25369..25670
                   /note="AluX repeat: matches 1..304 of consensus"
repeat_region     26305..26423
                   /note="MIR repeat: matches 20..142 of consensus"
repeat_region     26810..27184
                   /note="12 repeat: matches 1591..2004 of consensus"
repeat_region     27930..28238
                   /note="AluX repeat: matches 2..312 of consensus"
repeat_region     28300..28612
                   /note="AluX repeat: matches 1..311 of consensus"
misc_feature      complement(29577..29949)
                   /note="match: GSS: Em:AQ084247"
repeat_region     31407..33048
                   /note="L1M4 repeat: matches 80..1764 of consensus"
repeat_region     33049..33348
                   /note="AluX repeat: matches 1..298 of consensus"
repeat_region     33349..33598
                   /note="L1M4 repeat: matches 1764..2005 of consensus"
repeat_region     33689..33770
                   /note="Alu repeat: matches 227..308 of consensus"
repeat_region     33908..34591
                   /note="L1M4 repeat: matches 2257..2599 of consensus"
repeat_region     34593..34725
                   /note="AluY repeat: matches 2..134 of consensus"
repeat_region     34759..35417
                   /note="L1M4 repeat: matches 2614..3307 of consensus"
repeat_region     35415..36185
                   /note="L1PA13 repeat: matches 5375..6155 of consensus"
repeat_region     36330..36635
                   /note="AluSg repeat: matches 2..307 of consensus"
repeat_region     36833..36896
                   /note="MER72 repeat: matches 541..606 of consensus"
repeat_region     37018..37351
                   /note="AluYb repeat: matches 1..299 of consensus"
repeat_region     37352..38311
                   /note="L1M4 repeat: matches 3727..4692 of consensus"
repeat_region     38312..38456
                   /note="AluY repeat: matches 123..269 of consensus"
repeat_region     38457..38542
```

```
repeat_region     /note="43 copies 2 mer aa 62% conserved"
repeat_region     38637..39470
                   /note="L1MD1 repeat: matches 5354..6224 of consensus"
repeat_region     39479..39699
                   /note="AluY repeat: matches 84..303 of consensus"
repeat_region     39928..40268
                   /note="L1M1 repeat: matches 5609..5957 of consensus"
repeat_region     40269..40548
                   /note="AluX repeat: matches 1..277 of consensus"
repeat_region     40829..40963
                   /note="MIR repeat: matches 94..248 of consensus"

Query Match      2.1%; Score 40.6; DB 9; Length 168108;
Best Local Similarity 57.5%; Pred.No.5;
Matches 73; Mismatch 0; Mismatches 54; Indels 0; Gaps 0;
```

```
QY 61 aaacattgaatactgctgagagagtgacatgcataaggactatccataac 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91971 AAGATATAGATCATTCGCCAGTACCAATGATATCCCTCTGCTCTCAATCACAR 91912
```

```
QY 121 tctcttgacattatgtgtagtctccattctctgtgtaattccctgagtc 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91911 CCTCTTTTTCCTCTATTCGTATTCATTCCTGATTTTCTTTCCTTATCTGTGTG 91852
```

```
QY 181 tgtagtg 187
    || |||
Db 91851 TGCATGT 91845
```

```
RESULT 12
AL513187          122032 bp   DNA      linear   PRI 11-AUG-2001
LOCUS             Human DNA sequence from clone RP11-436H6 on chromosome 1, complete
DEFINITION
ACCESSION         AL513187
VERSION           AL513187.6   GI:15149851
KEYWORDS          HTG.
SOURCE            human.
ORGANISM          Homo sapiens
                  Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE         1 (bases 1 to 122032)
AUTHORS           Wallis,J.
TITLE             Direct Submission
JOURNAL           Submitted (11-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
                  CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
                  requests: clonequest@sanger.ac.uk
```

## COMMENT

On Aug 13, 2001 this sequence version replaced gi:14160999. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch1> RP11-436H6 is from the library RPc1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

\* NOTE: This is a 'working draft' sequence.

RESULT	14
AL590427	
LOCUS	213692 bp DNA linear HTG 01-FEB-2002
DEFINITION	Homo sapiens chromosome 1 clone RP11-336N10, *** SEQUENCING IN
ACCESSION	AL590427
VERSION	AL590427.12 GI:18491341
KEYWORDS	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE	1 (sites)
JOURNAL	Walls, J.
COMMENT	Direct Submission Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humanyes@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Feb 4, 2002 this sequence version replaced gi:117902916.

--- 020 bases; agalose-1p  
---  
---  
--- average; 10.000



Tue Aug 20 08:07:59 2002

us-09-662-812-1\_1.rge

Page 18

Oy 1696 ag 1697  
Db 221 RS 220

Search completed: August 20, 2002, 01:14:17  
Job time: 6607 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:24:10 ; Search time 2367.23 Seconds  
(without alignments)  
16858.060 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907

Sequence: 1 gtggtctgatttgaaaaag.....ctcatcaagagctgtgaa 1907

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No. Query Match Length DB ID Description

1	1907	100.0	1907	6	AX100528	Sequence
2	1907	100.0	8268	1	AE002234	Chlamydia
3	1907	100.0	16448	1	AE001587	Chlamydia
4	1707	89.5	1707	6	AX349527	Sequence
5	561.4	29.4	11627	1	AE001308	Chlamydia
6	511.8	26.8	11765	1	AE002331	Chlamydia
7	84	4.4	2088	6	AX349569	Sequence
8	43	2.3	7892	6	AX348443	Sequence
9	41.6	2.2	1141	6	AX083744	Sequence
10	40.6	2.1	128866	2	AP004050	Arabidops
11	40.6	2.1	168108	9	AL445687	Human DNA
12	40.4	2.1	122032	9	AL513187	Human DNA
13	40.4	2.1	161067	2	AL607133	Human DNA
14	40.4	2.1	213692	2	AL590427	Homo sapi
15	40	2.1	1141	6	AX083744	Sequence
16	39.6	2.1	22601	8	AB022222	Arabidops
17	39.6	2.1	69988	2	AC100581	Mus muscu
18	39.4	2.1	54585	2	AC100191	Mus muscu
19	39.4	2.1	59564	2	AC102477	Mus muscu
20	39.4	2.1	62396	2	AC100906	Mus muscu
21	39.4	2.1	165127	2	AC103172	Rattus no
22	39.2	2.1	1976	6	AX136235	Sequence
23	39.2	2.1	2311	9	AK027438	Homo sapi
24	39.2	2.1	2870	9	BC020959	Homo sapi
25	39.2	2.1	3389	9	AF160213	Homo sapi
26	39.2	2.1	3592	9	AF269150	Homo sapi
27	39.2	2.1	145593	9	AC069523	Homo sapi
28	39.2	2.1	157943	9	AC016938	Homo sapi
29	39.2	2.1	162820	9	AL138765	Human DNA
30	39	2.0	114314	9	AL359494	Human DNA
31	39	2.0	143840	2	AC091542	Felis cat
32	39	2.0	166602	2	AC022899	Homo sapi
33	39	2.0	169792	2	AC091508	Felis cat
34	39	2.0	180832	2	AC044828	Homo sapi
35	39	2.0	212310	2	AC008679	Homo sapi
36	38.8	2.0	24186	2	AC108968	Rattus no
37	38.6	2.0	92250	9	AC026693	Homo sapi
38	38.6	2.0	125020	9	AF429315	Homo sapi
39	38.6	2.0	156248	9	AC027322	Homo sapi
40	38.6	2.0	235564	2	AC022062	Mus muscu
41	38.4	2.0	11323	3	AF067951	Caenorhab
42	38.4	2.0	104972	9	AC025176	Homo sapi
43	38.4	2.0	152862	2	AL662968	Oryza sat
44	38.4	2.0	177838	2	AL606690	Oryza sat
45	38.4	2.0	178113	2	AC099516	Homo sapi

ALIGNMENTS

RESULT 1  
AX100528 LOCUS AX100528 1907 bp DNA linear PAT 10-APR-2001  
DEFINITION Sequence 1 from Patent WO0121804.  
ACCESSION AX100528  
VERSION AX100528.1 GI:13619532  
KEYWORDS  
SOURCE  
ORGANISM Chlamydia pneumoniae.  
REFERENCE Chlamydia pneumoniae  
1 (bases 1 to 1907)  
AUTHORS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
TITLE I chlamydia /I antigens and corresponding dna fragments and uses  
JOURNAL thereof  
Patent: WO 0121804-A 1 29-MAR-2001;  
Aventis  
FEATURES  
source location/Qualifiers  
1. 1907  
/organism="Chlamydia pneumoniae"  
/db\_xref="taxon:83558"  
101. 1807  
/note="unnamed protein product"  
/codon\_start=1

CDS

```
/transl_table=1
/protein_18="CAC36452.1"
/db_xref="GI:13619533"
/translation="MGLFHLTFGLILCSLPISLVAKPESVGHKILYISTQOAL
ATYLEADLAYDHDHDFVYLKIGEDLYLKOSIHSSPPRKSTIIAGLAGSEADLVS
QAMETADPLDOLLVLSAVSHGLKTSDDLKALISVPVRLAARLAKRKTRY
DHLSPFHKLPDEFIOCLSAATFRLFTESENAVYRDLLAKSKARSFTALDIEYD
KRLPTLRNLITSAPDQDAITLALGKLKDGOSTYNKLOLPDVDTLAAQAL
ALGEEDLPIVKQALEERPRALYRHPSEIGIPALPFLKTNSEKSNVLA
LLEIGDTPKLEYITRERLYOPHYNETLALSFSGRTLOMKRVNIIVPODEREL
LSTTRGLEEQILFRLPKKAYLPCIKYLLASQKQATLAIISLTSVSHOEDLL
FOAKLPEGLIRAYADAIYNLKDPEKRSKLDVAKPFLKLIQELFLVDENQPHS
MPYLRVQVTPESRKLMDILLETATSKSSDILLLQIMBGAKNPVLGLIKI
VE"
BASE COUNT      542 a      456 c      380 g      529 t
ORIGIN
Query Match      100.0%; Score 1907; DB 6; Length 1907;
Best Local Similarity 100.0%; Pled. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtggcttgatttgaaaaaggccatgagatggttataatggtcaagtcctccatcc 60
Db 1 GTGGCTTGATTGTGAAAAAGGCCATGGATGTTTATATGTTCAAGTCTCCCTATCC 60
QY 61 aaacatgaaactctgtagagaggttgaacatcgatctatggaactatccatctaac 120
Db 61 AAACATGAAATACTTGTGTAGAGGAGTGAACATCGATCTATGGACATTTCCATCTAAC 120
QY 121 tctcttggaactttatggtgtagctctccatctctcttcttgtaattccctgaatc 180
Db 121 TCTCTTGGAAGTTTATTTATTTGTAGTCTTCCATTTCTCTTGTTGTTAAATTCCTGAGTGC 180
QY 181 ttagagtcataagatcccttitalaagtaacgaatctacacagagccttagacaata 240
Db 181 TGTAGTGTATAGATCTTTATATAGTAGCAATCTACACAGCAGCGCTTAGCAACATA 240
QY 241 tctggaagctctagatgcctcagtgatcatgactctctgcttctaagaaaaatcgaga 300
Db 241 TCTGGAAGCTCTAGATGCTCGGTGATCATGACTTCTCTTTTAAAGAAAAATCGGAGA 300
QY 301 agactatctcaagcaaaagcatccactctcagatccggaacttagaaaaagcacatcat 360
Db 301 AGACTATCTCAAGCAAAAGCATCCACTCTCGATCCGCAAACTAGAAAAACACCATCAT 360
QY 361 tggagcagagccttgcggagatcttcaagaagccttggacgtgtcttcccaagatggaaac 420
Db 361 TGGAGCAGAGCCTTGGCGGAGTCTCAAGAACCTTGGAGTGCTCTCCCAAGCTATGGAAAC 420
QY 421 tgcgaagccctctcagcagcagctagtgtttatcgggcagctcgaagacatcttggaaac 480
Db 421 TGCAGACCCCTCGAGAGAGCTACTGGTTTTATCGCAGTCTCAGAGCATTTGGGAAAC 480
QY 481 ttccgaagcactactgttttaagcttttagacatcccatctccttctcactcggctagaagc 540
Db 481 TTCTGAGAGACTTACTGTTTAAAGCTTTTAGCATCTCCATCTCTGATCCGCTTAGAGAC 540
QY 541 cgcctatagacttgcataatttgaagaaacataaagtcattgatcatctacatcttccat 600
Db 541 CGCTATAGACTTGCTAATTTGAAGAACATAAGTCAATTGATCATCTACTTCTTTTCAT 600
QY 601 tcataagcttccggaagaataatcaatgcatacttgcggcaatatctcctaagcttggagac 660
Db 601 TCATAAGCTTCCGGAAGAAATCCAAATGCTATTCGCGCAATATTTCTTACGCTTGGAGAC 660
QY 661 tgaagaactcgaatgctataatctcggaatctcttagctgcgaagaagaagcgcgatctggag 720
Db 661 TGAAGAATCTGATGCTTATATTCGGGATCTCTTAGCTGCGCAAGAAAGCGGATTCGGAG 720
QY 721 tgcacagacttgcagatctcggaataacaaacaaacgcttcttcggaacttagaa 780
Db 721 TGCACAGACTTGTGAGATCGGAGAAATACCAAAACAAACGCTTCTTCCGACACTTAGGAA 780

QY 781 ttctgtaacgagtcgctcctccaagatcaagaagcatcttcttagctttagggagct 840
Db 781 TTTGTAACGAGTGGGTCTCTCAAGATCAACAGCATATCTTTATGCTTTAGGGAAGCT 840
QY 841 taaggaatggtcagagctactacataataaaaaaagaatgcgaagcctgtagtctgagt 900
Db 841 TAAGGATGGTCAAGACTACTCAATATATAAAAAAGCAATTTGCGAAGCCGATGTGATGT 900
QY 901 caacttaagaagagctcaagctttaaagcttcttgagggaagaagagcgctccctccgt 960
Db 901 CACTTTAGCAGAGCTCAAGCTTTAATGCTTTTGGGGAAGAGAGAGAGCGTCTTCCGCT 960
QY 961 gataaaaaagcaagcacttgaagagcgagcctcgaagccctgtatgtccttaagcagcatcac 1020
Db 961 GATAAAAAAGCAAGCACTTGAAGAGCGGCGCTGAGCCCGTATGCTTAGCGCATCTAC 1020
QY 1021 ctctgaagataggagatccgatctgccttgcgatatctcctaaactaagaacaggaagc 1080
Db 1021 CTCTGAGATAGGAGATTCGATTCGCCGATTCCTTAAAACTAAGAACAGCGAAGC 1080
QY 1081 caagttgaatgtagctttagctctctcttagagtgtagtgtagacacccctaaactatgga 1140
Db 1081 CAAGTTGAATGATGATCTTAGCTCTCTTGAAGTTAGGGTGTGACCCCTTAAACTACTGGA 1140
QY 1141 atacattacggaaggtctgtcccaacacatataatgagactcagacttgaagttctc 1200
Db 1141 ATACATTAACGGAAGGCTTGTCACCAACACATTAATGAGACTGATGAGTTCTC 1200
QY 1201 taagggcgctacttacaacaaatctgaaagcggtgaacatcatagctccctaaagatcccca 1260
Db 1201 TAAAGGCGCTACTTACAAATTTGAAAGCGGTTGAACATCATAGTCCCTCAAGATCCCA 1260
QY 1261 ggaagaggaaggttgcctccacaacccgaggtcttgaagagcagatcccttagcttct 1320
Db 1261 GGAAGGGAAGGTTGCTCTCCACAAACCGAGGCTTTAAGAGCAGATCTTACGTTTCT 1320
QY 1321 ctccgcctacctaagaagatctacccctgttattataagcttcttgaggagtcagaa 1380
Db 1321 CTCCGCTTACCTTAAGAAAGCTTACCTCCGATTTATTAAGCTTTGGCGAGTCAGAA 1380
QY 1381 aactcagcttgcacactctgagatctcttctttaaagtaacacccctcaacatcgaagcctt 1440
Db 1381 AACTCAGCTTGCACACTACTGTGGATTTCTTTTAAAGTCAACCTCACATYCGAGAAAGCC 1440
QY 1441 agatctacttccaagcttgcgaagctcctgcggaaacctatccatcgcgctatgcaga 1500
Db 1441 AGATCTACTTTTCCAAGCTGCGAAAGCTTCTGGAAGAACTTATCATCCGCTTATGCAGA 1500
QY 1501 tcttgctattataatctcaaccaaagatcctgnaaaaaaacgcttctcctcaatgatalgc 1560
Db 1501 TCTTGCTATTATTAATCTCACCAAAAGATCTGAAAAAAAAGTTCCTCCATGATTAATGC 1560
QY 1561 aaaaaagctaatcaagaagaacctgttatctgtgagcagggaaaaacaaagaccccatcc 1620
Db 1561 AAAAAAGCTAATCAAGGAACCTTTGTTATTTGTGACACGGAAAAACCAAAACCCCATCC 1620
QY 1621 caagcatccctatcagttacgttaccgaagcctgcgaagaagcgttagaagctcagttgga 1680
Db 1621 CAGCATGCCCTATCTAGTTATCAAGTTCACCCGAAAGCCGTCAGAAAGCTCATGTTGGA 1680
QY 1681 tatctagagaacttagccacactcgaagctctccgaagataatccgttatatgatacaact 1740
Db 1681 TATCTAGAGACACTAGCACCACTCGAAGTCTCGAAGATATCGGTTATATGATACAACT 1740
QY 1741 gatgacggaaggaatgcaaaaaaatcttccagatctcctgcgaggtctactcataaaaaatgtc 1800
Db 1741 GATGACGGAAGGAGATCAAAAAATTTTCCAGTCTTGCAGGCTTACTCATTAATAAATTTGT 1800
QY 1801 ggaataaacccaactcagttctttagaaggttcttcttattcttagcttccctctgttc 1860
Db 1801 GGAATTAACCCCACTACGCTTATGAAGGTTGCTTATTTCTTAGCTTCCCTTTGTTC 1860
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:59:25 ; Search time 226.66 Seconds  
(without alignments)  
14445.230 Million cell updates/sec

Title: US-09-662-812-1  
Perfect score: 1907  
Sequence: 1 gtggctgatttgaaaag.....ctcatcaagagcgtgtgaa 1907

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:\*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1907	100.0	1907	22	AAD03024
2	39.2	2.1	1867	22	AAH46935
3	39.2	2.1	1976	22	AAK93822
4	39.2	2.1	2299	22	AAK94454
5	39.2	2.1	2311	22	AAH13994
6	39.2	2.1	2864	22	AAK94275
7	39.2	2.1	3076	22	AAH46951
8	39.2	2.1	3370	21	AAH16630
9	38.2	2.0	38771	22	AAK66361

C	10	38.2	2.0	38771	22	AAK68883	Human immune/haema
C	11	38.2	2.0	38771	22	AAK69604	Human immune/haema
C	12	38.2	2.0	38771	22	AAK71499	Human immune/haema
C	13	38.2	2.0	38771	22	AAK72925	Human immune/haema
C	14	38.2	2.0	38771	22	AAK75870	Human immune/haema
C	15	38.2	2.0	38771	22	AAK80325	Human immune/haema
C	16	38.2	2.0	38771	22	AAK83538	Human immune/haema
C	17	37.2	2.0	932	21	AAK16080	Human prostate can
C	18	36.8	1.9	4940	16	AAO91286	ORC1 gene encoding
C	19	36.8	1.9	4940	18	AAK73279	S. cerevisiae orig
C	20	36.4	1.9	633	7	AAH60340	Sequence encoding
C	21	36	1.9	549	22	AAK44753	Human contig polyn
C	22	36	1.9	1180	22	AAH70004	Tomato glutamate d
C	23	36	1.9	2536	22	AAK64216	Human secreted pro
C	24	35.8	1.9	467	22	AAK93524	Human polynucleoti
C	25	35.6	1.9	336	22	AAH12496	Human cDNA clone (
C	26	35.6	1.9	2459	22	AAH16543	Human cDNA sequenc
C	27	35.4	1.9	21748	23	ABL02520	Drosophila melanog
C	28	35.2	1.8	482	22	AAH94479	Human neuroblastom
C	29	35.2	1.8	2379	22	AAH61120	Human polynucleoti
C	30	35.2	1.8	2387	23	ABL23408	Drosophila melanog
C	31	35.2	1.8	2413	22	AAH59334	Human polynucleoti
C	32	35.2	1.8	2498	21	AAK23778	Human fetal brain
C	33	35.2	1.8	3466	22	AAH16111	Human cDNA sequenc
C	34	35.2	1.8	4222	23	ABL20216	Drosophila melanog
C	35	35.2	1.8	4755	23	ABL01896	Drosophila melanog
C	36	35.2	1.8	6021	23	ABL14540	Drosophila melanog
C	37	35.2	1.8	6692	23	ABL14942	Drosophila melanog
C	38	35.2	1.8	6692	23	ABL20236	Drosophila melanog
C	39	35	1.8	6692	23	ABL07906	Drosophila melanog
C	40	35	1.8	4640	23	ABL08094	Drosophila melanog
C	41	34.8	1.8	633	7	AAH60341	Sequence encoding
C	42	34.8	1.8	1201	16	AAO98905	Coho Salmon growth
C	43	34.8	1.8	33962	21	AAH81528	N. meningitidis pa
C	44	34.8	1.8	34980	21	AAK21609	Neisseria meningit
C	45	34.8	1.8	34980	21	AAK21610	Neisseria meningit

#### ALIGNMENTS

RESULT 1

AAD03024

ID AAD03024 standard; DNA; 1907 BP.

AC AAD03024;

DT 13-JUN-2001 (first entry)

DE Chlamydia pneumoniae outer membrane protein (OMP) gene.

KM Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;

KW antibody; medicament; Chlamydia infection; ds.

OS Chlamydia pneumoniae.

XX

FT Key Location/Qualifiers

FT CDS 101..1807

FT /tag= a

FT /product= "C. pneumoniae outer membrane protein (OMP)"

PD WO200121804-A1.

PD 29-MAR-2001.

PF 15-SEP-2000; 2000WO-CA01088.

PR 20-SEP-1999; 99US-0154652.

XX (AVET) AVENTIS PASTEUR LTD.

XX Murdin AD, Oomen RP, Wang J, Dunn P;

PI

DR WP1: 2001-244939/25.  
DR P-PSDB: AAY72972.  
XX  
PT Novel Chlamydia pneumoniae outer membrane protein and polynucleotides  
PT encoding them, useful as components of vaccines for treating Chlamydia  
PT infections, and for detecting Chlamydia infections in the body fluids  
PT of mammals.  
XX  
PS Claim 2: Fig 1; 82pp; English.  
XX  
CC The present sequence is a DNA encoding Chlamydia pneumoniae OMP (outer  
CC membrane protein). OMP is useful in the production of vaccines,  
CC antibiotics and pharmaceutical compositions which are useful for treating  
CC or preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or  
CC C. pecorum) infections. They are also useful as diagnostic reagents for  
CC detecting Chlamydia infection which involves assaying a body fluid  
CC of a mammal to be tested for the components. The OMP vaccine is  
CC useful in the preparation of a medicament for preventing and/or  
CC treating Chlamydia infection. The primers derived from OMP gene are  
CC also useful for detecting and/or identifying Chlamydia in a biological  
CC material. OMP antibodies are also useful as reagents for purifying OMP  
CC from a biological sample which involves carrying out antibody-based  
CC affinity chromatography with the biological sample. OMP gene is  
CC also useful in gene therapy.

XX  
SQ Sequence 1907 BP; 542 A; 456 C; 380 G; 529 T; 0 other;

Query Match 100.0%; Score 1907; DB 22; Length 1907;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgcttgatttgaaaggctcagatgatttataatgctcaagcttcctacc 60  
DB 1 gtgcttgatttgaaaggctcagatgatttataatgctcaagcttcctacc 60  
QY 61 aaacatgaaatctgctagagagttgaaacatcgatcctatggacatcctaac 120  
DB 61 aaacatgaaatctgctagagagttgaaacatcgatcctatggacatcctaac 120  
QY 121 tctcttgagatttattgtgtatcttccattctcttctgtgttaattccctgagtc 180  
DB 121 tctcttgagatttattgtgtatcttccattctcttctgtgttaattccctgagtc 180  
QY 181 tgtagtcataagaatcctttatagaatgacaaatcacaagaagccttaagaacata 240  
DB 181 tgtagtcataagaatcctttatagaatgacaaatcacaagaagccttaagaacata 240  
QY 241 tctggaagctctagatgctacggttgatcagtaactcttctgtttaaagaanaacgagaga 300  
DB 241 tctggaagctctagatgctacggttgatcagtaactcttctgtttaaagaanaacgagaga 300  
QY 301 agatatactcaagaagaagcatcactcctcagatcgcgaacatagaaagacacatcat 360  
DB 301 agatatactcaagaagaagcatcactcctcagatcgcgaacatagaaagacacatcat 360  
QY 361 tggagcagcctgagcggatcttcagaagccttgagtgctctcccaagcttagaagac 420  
DB 361 tggagcagcctgagcggatcttcagaagccttgagtgctctcccaagcttagaagac 420  
QY 421 tggagacccctgagcagctactgtttatcgagcagcttcagagacatttgggaaac 480  
DB 421 tggagacccctgagcagctactgtttatcgagcagcttcagagacatttgggaaac 480  
QY 481 tcttgacgactactgtttaaagcttagacatccctacatcctgctacgttagagac 540  
DB 481 tcttgacgactactgtttaaagcttagacatccctacatcctgctacgttagagac 540  
QY 541 cgcctatagacttgtaatttgaagaacataaagtaagtaagtaagtaagtaagtaag 600  
DB 541 cgcctatagacttgtaatttgaagaacataaagtaagtaagtaagtaagtaagtaag 600  
QY 601 tcatagactcccgagaagaatcacaatgacatctgagcaatatctcctagccttgagac 660

DB 601 tcatagactcccgagaagaatcacaatgacatctgagcaatatctcctagccttgagac 660  
QY 661 tgaagaatctgattatattcggatctcttagctgcgaagaagaagcagctcggag 720  
DB 661 tgaagaatctgattatattcggatctcttagctgcgaagaagaagcagctcggag 720  
QY 721 tgcacagcttgcagatgcggaataccacaagaacgcttcttcgacacattagana 780  
DB 721 tgcacagcttgcagatgcggaataccacaagaacgcttcttcgacacattagana 780  
QY 781 ttgttaacgagctgctcctcagaatcaagaacatcttcttctgtcttgggaagct 840  
DB 781 ttgttaacgagctgctcctcagaatcaagaacatcttcttctgtcttgggaagct 840  
QY 841 taagatgctcagagctactacaataataaaagaatgcagaagcctgagtgtatgt 900  
DB 841 taagatgctcagagctactacaataataaaagaatgcagaagcctgagtgtatgt 900  
QY 901 cactttagcagcagctcaagcttataatgcttggggaagaagaagccttcctcgt 960  
DB 901 cactttagcagcagctcaagcttataatgcttggggaagaagaagccttcctcgt 960  
QY 961 gataaaagcagaagcacttgagagcgctcgcagcctctgatatgcttagcgacatcacc 1020  
DB 961 gataaaagcagaagcacttgagagcgctcgcagcctctgatatgcttagcgacatcacc 1020  
QY 1021 ctctgagatagagatccgattgccttcgcgatatctcctaaanaaataagaacagcgaagc 1080  
DB 1021 ctctgagatagagatccgattgccttcgcgatatctcctaaanaaataagaacagcgaagc 1080  
QY 1081 caagttgaatgtagctttagctctcttagagttgaggtgtgacacccctaactctgta 1140  
DB 1081 caagttgaatgtagctttagctctcttagagttgaggtgtgacacccctaactctgta 1140  
QY 1141 atacattccgaagaagcctgacacacacatataaagaagccttagcctgaattctc 1200  
DB 1141 atacattccgaagaagcctgacacacacatataaagaagccttagcctgaattctc 1200  
QY 1201 taagggcgtaacttatacaanaattggaagcgggtgaacatcctagctcccaagaatcccca 1260  
DB 1201 taagggcgtaacttatacaanaattggaagcgggtgaacatcctagctcccaagaatcccca 1260  
QY 1261 ggaagagggaaggtgtgctctccacaacccgaagctcttgaagaagcagatccttaagcttct 1320  
DB 1261 ggaagagggaaggtgtgctctccacaacccgaagctcttgaagaagcagatccttaagcttct 1320  
QY 1321 ctccgcctactcaagaagcttaccctccctgattatagaagcttggggaagtaagaa 1380  
DB 1321 ctccgcctactcaagaagcttaccctccctgattatagaagcttggggaagtaagaa 1380  
QY 1381 aactcaagcttgcactactgagattcttctttaaagtaagcctcacaatcaggaagcctt 1440  
DB 1381 aactcaagcttgcactactgagattcttctttaaagtaagcctcacaatcaggaagcctt 1440  
QY 1441 agatctacttctcaagctgcgaagcttctctggaacatcatccgcgcttagagaga 1500  
DB 1441 agatctacttctcaagctgcgaagcttctctggaacatcatccgcgcttagagaga 1500  
QY 1501 tcttgctatttataatctccacaagaatcctggaanaaaagcttctcctcatgattatgc 1560  
DB 1501 tcttgctatttataatctccacaagaatcctggaanaaaagcttctcctcatgattatgc 1560  
QY 1561 aaaaaagctatcacaagaacgtgtattgtggaacggaagaacacaagaaccccatcc 1620  
DB 1561 aaaaaagctatcacaagaacgtgtattgtggaacggaagaacacaagaaccccatcc 1620  
QY 1621 cagcatgcccctatcactaagttatcaggtcaccgagaagcgttagcagaagctatgttga 1680  
DB 1621 cagcatgcccctatcactaagttatcaggtcaccgagaagcgttagcagaagctatgttga 1680  
QY 1681 tattctagagaacatgacacctcgaagcttctcgaagaatatccgttattatgataaact 1740

Db	1661	tattctagaagacactcagccacctcgaaagctctccgaagatcccgctttattgtataact	1740
QY	1741	gatgcaggaagagatgtcaaaaaatttccagcttcttgcaaggcttactcataaaattgt	1800
Db	1741	gatgcaggaagagatgtcaaaaaatttccagcttcttgcaaggcttactcataaaattgt	1800
QY	1801	ggaggaaccccaactaagctcttafgaaacgttgccttcttaattctagcttccttgctc	1860
Db	1801	ggaggaaccccaactaagctcttafgaaacgttgccttcttaattctagcttccttgctc	1860
QY	1861	ttatgggttcctcagctatgcttttgaccatcaagaagctgtgaaa	1907
Db	1861	ttatgggttcctcagctatgcttttgaccatcaagaagctgtgaaa	1907
RESULT 2			
AAH46935/c			
ID	AAH46935	standard; CDNA: 1867 BP.	
XX	AAH46935:		
AC			
XX			
DT	25-SEP-2001	(first entry)	
XX			
DE		Human secreted protein encoding CDNA (clone Id HB06C11).	
XX			
KW		Secreted protein; immunosuppressive; antiarthritic; antiinflammatory;	
KW		antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;	
KW		neotropic; neuroprotective; antibacterial; virucide; fungicide; human;	
KM		ophthalmological; gene therapy; ss.	
XX			
OS		Homo sapiens.	
XX			
PN	WO200155430-A1.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01431.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	12-SEP-2000; 2000US-0231968.		
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Moore CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;		
PI	Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;		
PI	Ni J, Ruben SM, Barrash SC.		
XX			
DR	WPI: 2001-476220/51.		
DR	P-PSDB: AAB85525.		
XX			
PT	17 isolated nucleic acid molecules encoding human secreted proteins,		
XX	used to preventing, treating or ameliorating a medical condition		
XX			
XX	Claim 1; Page 414-415; 482pp; English.		

CC	supporting cell culture of primary tissues, to regenerate tissues and in
CC	chemotaxis. The polypeptides can also be used as a food additive or
CC	preservative to increase or decrease storage capabilities. The present
CC	sequence represents a human secreted protein encoding CDNA.
xx	
SQ	Sequence 1867 BP; 542 A; 315 C; 347 G; 663 T; 0 other;
Query Match	2.1%; Score 39.2; DB 22; Length 1867;
Best Local Similarity	47.9%; Pred. No. 0.62;
Matches 113; Conservative	0; Mismatches 123; Indels 0; Gaps 0
OY	14 gaaaaggtccatgcatgtgtttataatgltcaaggtccctatccaacatggaata 73
Db	1420 GAAAAAGATGATTCACCTTTAGACATCATCTTAAAGTAGTTATCTCAAAATG 1361
OY	74 ctgttagagaggttgaacatcgatctatgggacattatccatactctcttggaact 133
Db	1360 TTTAATTAATAACAAGTATCTTCCACATTTAAACCTTTGGCTTTAACTGTACAGTAATT 1301
OY	134 ttatgtgtgagcttcccatctctctgttgcctaaatccctgagctgtgaagtcataag 193
Db	1300 GCATGTGTAGAGAGTACACTTCTGTCTTCAAACTGTATCTTCTTTGGATGGAATTAAGATG 1241
OY	194 alccttataataagtaagcaatctacacagcagcgcttagcaacatatctggaagc 249
Db	1240 TAACGTATATAGTTTAAAGATTAATAATGGAAGTGTGTCACACTGAATGACAGC 1185
RESULT 3	
AAF93822/c	
ID	AAF93822 standard; CDNA; 1976 BP.
xx	
AC	AAF93822;
xx	
DT	23-MAY-2001 (first entry)
xx	
DE	Human cDNA encoding a membrane or secretory protein clone PSEC0143.
xx	
KW	Human; secretory protein; membrane protein; vaccine; gene therapy;
KX	rheumatoid arthritis; diabetes; ss.
xx	
OS	Homo sapiens.
xx	
PN	EP1067182-A2.
xx	
PD	10-JAN-2001.
xx	
PF	07-JUL-2000; 2000EP-0114090.
xx	
PR	08-JUL-1999; 99JP-0194179.
PR	11-JAN-2000; 2000JP-0118775.
PR	02-MAY-2000; 2000JP-0183766.
xx	
PA	(HELI-) HELIX RES INST.
xx	
PI	Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
xx	
DR	WPI; 2001-093989/11.
xx	
DR	P-PSDB; AAB88395.
xx	
PT	Nucleic acids encoding secretory proteins/membrane proteins, useful in
xx	gene therapy or as candidate target molecules in drug development -
xx	
PS	Claim 1; SEQ ID 157; 609pp + CD ROM; English.
xx	
CC	This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC	which encode human secretory or membrane proteins represented by
CC	AAB88317 - AAB88419. Included in the invention are primers
CC	AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC	CDNA sequences of the invention. The invention also includes methods for
CC	the production of antibodies directed against the proteins, and cDNA
CC	sequences, which can be used in vaccines. The polynucleotide sequences

CC can be used in gene therapy. The polynucleotide sequences and the  
CC proteins they encode may be used in the prevention, treatment and  
CC diagnosis of diseases associated with inappropriate secretory  
CC protein/membrane protein expression. The nucleic acids and complementary  
CC sequences may also be used as DNA probes in diagnostic assays  
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
CC presence of similar nucleic acid sequences in samples. They may also be  
CC used to study the expression and function of secretory proteins/membrane  
CC polypeptides and their role in metabolism. The polypeptides may be used  
CC as antigens in the production of antibodies against them and in assays to  
CC identify modulators (agonists and antagonists) of expression and  
CC activity. The antibodies and antagonists may also be used as therapeutic  
CC agents to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of the  
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
CC (ELISA)). Examples of diseases which may be treated include rheumatoid  
CC arthritis and diabetes.  
CC  
XX  
SQ Sequence 1976 BP; 595 A; 314 C; 371 G; 696 T; 0 other;

Query Match 2.1%; Score 39.2; DB 22; Length 1976;  
Best Local Similarity 47.9%; Pred. No. 0.65;  
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 14 gaaaaaggtccatgagtggtttataatggtcaagggtccctccatccaacattgaata 73  
DB 1231 GAAAAAAGATATACCTTTAGACATTCAATAAATGTATTAATCAAAATG 1172  
QY 74 ctctgtagagagtggaacatcgatcctatggaacatccatccatcctcttgactt 133  
DB 1171 TTTAATTAATAAACAGTATCTTCTCCATTAAACCTTCTTCTTAACGTGACAGTAAT 1112  
QY 134 ttatgtgtagtctccatctctctgtgtctaattccctgagtcgtgagtcataag 193  
DB 1111 GCATTGTGAGAGTACACTTCTGCTTCAAACTGATCTTCTTGATGAGATTAAGATG 1052  
QY 194 atcctttataaagtagcgaatcctacacagcagccttagaacaacatctggaagc 249  
DB 1051 TAACGTATAGTTTAAAGATAAATAATGGGAAGTTGGTCCCAACTAAGATGACAGC 996

RESULT 4  
AAK94454/C  
ID AAK94454 standard; CDNA: 2299 BP.  
XX  
AC AAK94454;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human full-length cDNA, SEQ ID NO: 3256.  
XX  
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PE 07-JUL-2000; 2000EP-0114089.  
XX  
PR 08-JUL-1999; 99JP-0194486.  
PR 11-JAN-2000; 2000JP-0118774.  
PR 02-MAY-2000; 2000JP-0183765.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
XX WPI: 2001-524255/58.  
DR P-PSDB: AAM93524.

XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their  
PT use in genetic manipulation.  
XX  
PS Claim 8; SEQ ID NO 3256; 1380bp + sequence listing; English.  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a full length  
CC human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
CC  
XX  
SQ Sequence 2299 BP; 649 A; 380 C; 457 G; 813 T; 0 other;

Query Match 2.1%; Score 39.2; DB 22; Length 2299;  
Best Local Similarity 47.9%; Pred. No. 0.7;  
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 14 gaaaaaggtccatgagtggtttataatggtcaagggtccctccatccaacattgaata 73  
DB 1867 GAAAAAAGATATACCTTTAGACATTCAATAAATGTATTAATCAAAATG 1808  
QY 74 ctctgtagagagtggaacatcgatcctatggaacatccatccatcctcttgactt 133  
DB 1807 TTTAATTAATAAACAGTATCTTCTCCATTAAACCTTCTTCTTAACGTGACAGTAAT 1748  
QY 134 ttatgtgtagtctccatctctctgtgtctaattccctgagtcgtgagtcataag 193  
DB 1747 GCATTGTGAGAGTACACTTCTGCTTCAAACTGATCTTCTTGATGAGATTAAGATG 1688  
QY 194 atcctttataaagtagcgaatcctacacagcagccttagaacaacatctggaagc 249  
DB 1687 TAACGTATAGTTTAAAGATAAATAATGGGAAGTTGGTCCCAACTAAGATGACAGC 1632

RESULT 5  
AAH13994/C  
ID AAH13994 standard; CDNA: 2311 BP.  
XX  
AC AAH13994;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:11070.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
KW Homo sapiens.  
XX  
OS  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PE 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.  
XX PT primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX PS Claim 8; SEQ ID 11070; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 2311 BP; 655 A; 382 C; 458 G; 816 T; 0 other;

Query Match 2.1%; Score 39.2; DB 22; Length 2311;  
Best Local Similarity 47.9%; Pred. No. 0.7;  
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaggtccatgagatggtttataatgltcgaaggtccctatccaacatgaaata 73  
DB 1879 GAAAGAAAGATATCATCTTTAGACATTCAGTTAAATGTAATTAATCTCAAAATG 1820  
DB 74 ctgtcagagaggttgacatcgatcgaaggtccatccatcctcttggact 133  
DB 1819 TTTAATATAAACAAGATATCTTCCATTAACTTTGCTTCTAACTGTACAGTAATTT 1760  
QY 134 ttattgtgtagcttcccatctctgtgtcctaattccctgagctgtgagtcataag 193  
DB 1759 GCATTGTAGAGAGTACACTCTGTCTTCAAACTGTATCTTTGGATGGAAATTAAGATG 1700  
QY 194 atccctataataagtaagcaatctacacagcaggtccttagcaacatactcggaaagc 249  
DB 1699 TAACTGTATAGTTTAAAGATTAATTAATGGGAAGTTGGTCCAACTAAGATGACAGC 1644

RESULT 6  
AAK94275/c  
ID AAK94275 standard; cDNA; 2864 BP.  
XX  
XX AC AAK94275;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX DE Human full-length cDNA, SEQ ID NO: 2908.  
XX  
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN EP1130094-A2.  
XX

PD 05-SEP-2001.  
XX  
XX 07-JUL-2000; 2000EP-0114089.  
XX  
XX 08-JUL-1999; 99JP-0194486.  
XX  
XX 11-JAN-2000; 2000JP-0118774.  
XX  
XX 02-MAY-2000; 2000JP-0183765.  
XX  
XX  
XX PA (HELI-) HELIX RES INST.  
XX  
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX WPI: 2001-524255/58.  
XX  
XX DR P-PSDB; AAM93355.  
XX  
XX PT 830 Primers useful for synthesizing full length cDNA clones and their  
XX use in genetic manipulation -  
XX  
XX PS Claim 8; SEQ ID NO 2908; 1380pp + sequence listing; English.  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesised by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a full length  
CC human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
XX  
SQ Sequence 2864 BP; 834 A; 463 C; 569 G; 998 T; 0 other;

Query Match 2.1%; Score 39.2; DB 22; Length 2864;  
Best Local Similarity 47.9%; Pred. No. 0.8;  
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 14 gaaaaggtccatgagatggtttataatgltcgaaggtccctatccaacatgaaata 73  
DB 2119 GAAAGAAAGATATCATCTTTAGACATTCAGTTAAATGTAATTAATCTCAAAATG 2060  
DB 74 ctgtcagagaggttgacatcgatcgaaggtccatccatcctcttggact 133  
DB 2059 TTTAATATAAACAAGATATCTTCCATTAACTTTGCTTCTAACTGTACAGTAATTT 2000  
QY 134 ttattgtgtagcttcccatctctgtgtcctaattccctgagctgtgagtcataag 193  
DB 1999 GCATTGTAGAGAGTACACTCTGTCTTCAAACTGTATCTTTGGATGGAAATTAAGATG 1940  
QY 194 atccctataataagtaagcaatctacacagcaggtccttagcaacatactcggaaagc 249  
DB 1939 TAACTGTATAGTTTAAAGATTAATTAATGGGAAGTTGGTCCAACTAAGATGACAGC 1884

RESULT 7  
AAH46951/c  
ID AAH46951 standard; cDNA; 3076 BP.  
XX  
XX AC AAH46951;  
XX  
XX DT 25-SEP-2001 (first entry)  
XX  
XX DE Human secreted protein encoding cDNA (clone id HBOEG11).  
XX  
XX KW Secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
XX antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
XX neurotropic; neuroprotective; antibacterial; virucide; fungicide; human;  
XX ophthalmological; gene therapy; ss.  
XX



O5	Homo sapiens.
XX	
PN	WO200155430-A1.
PD	
02-AUG-2001.	
17-JAN-2001; 2001WO-US01431.	
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	12-SEP-2000; 2000US-0231968.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
PI	Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
PI	Moore PA, Mel P, Ederer R, Duan DR, Shi Y, Choh GH, Fiscella M;
PI	Ni J, Ruben SM, Barash SC;
XX	
DR	WPI: 2001-476220/51.
P-PSTDB; AAB85541.	
17 Isolated nucleic acid molecules encoding human secreted proteins,	
used to preventing, treating or ameliorating a medical condition -	
Claim 1: Page 426-427; 482pp; English.	
The invention provides novel human secreted proteins and polynucleotides	
encoding them. The secreted proteins can be expressed by standard	
recombinant methodology. The secreted proteins and polynucleotides are	
used to prevent, treat or ameliorate a medical condition in e.g. humans,	
mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can	
also be used in diagnosing a pathological condition. The antibodies to	
the proteins can also be used in alleviating symptoms associated with the	
disorders and in diagnostic immunoassays e.g. radiimmunoassays or enzyme	
linked immunosorbent assays (ELISA). Disorders which are diagnosed or	
treated include autoimmune diseases e.g. rheumatoid arthritis,	
hyperproliferative disorders e.g. neoplasms of the breast or liver,	
cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders	
e.g. cerebral ischemia, angioneurosis, nervous system disorders e.g.	
Alzheimer's disease, infections caused by bacteria, viruses and fungi and	
ocular disorders e.g. corneal infection. The polypeptides can also be	
used to aid wound healing and epithelial cell proliferation, to prevent	
skin aging due to sunburn, to maintain organs before transplantation, for	
supporting cell culture of primary tissues, to regenerate tissues and in	
chemotaxis. The polypeptides can also be used as a food additive or	
preservative to increase or decrease storage capabilities. The present	
sequence represents a human secreted protein encoding cDNA.	
SQ Sequence 3076 BP; 940 A; 487 C; 617 G; 1028 T; 4 other:	
Query Match 2.1%; Score 39.2; DB 22; Length 3076;	
Best Local Similarity 47.9%; Pred. No. 0.83;	
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0	
OY 14 gaaagaagtcacgtgatgattataatgtcaaggctcccatccaacatlgaanaata 73	
Db 2570 GAAAGGATGATCTACTTGAACATTCACTTAATAAAGTAGTTATTCATCAACAANAAG 2511	
OY 74 ctgtcctaagagagtggaacaatcgatcactatggagacattccatcctaactccttggacct 133	
Db 2510 TTTAATTAATAAACAAAGTATCTTCATCATTTAAACACTTTGGTTCTTCACTACAGTAAAT 2451	
OY 134 ttatgttgttagctcccatctctctgtgttgtaaatcccgcagagtctagtgtcctaag 193	
Db 2450 GCATTGTGAGAGCTCACACTTCTGTCTCAAACTGATATCTTTTGATGGAATTAAGATG 2391	
OY 194 atccttatataagtcacgaatcctcacagcagcgcttagcacaacatctggaagc 249	
Db 2390 TAACGTATFAGTTTTAAAGATAATAATGGAAGTTGGTCCAAACTGAATGACAGC 2335	
RESULT 8	

ID	AAAI6630/c
ID	AAAI6630 standard; CDNA; 3370 BP.
XX	
AC	AAAI6630;
XX	
DT	16-JUN-2000 (first entry)
XX	
DE	Human secreted protein clone pk366_7 nucleotide sequence SEQ ID NO:25.
XX	
KW	Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW	antibacterial; antitumoral; cytostatic; antiinflammatory; dermatological;
KW	antidiabetic; antiasthmatic; antiarthritic; antineoplastic; protozoacide;
KW	antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW	infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW	connective tissue disease; multiple sclerosis; erythematosus;
KW	rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW	Gullain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW	insulin dependent diabetes mellitus; graft-versus-host-disease;
KW	autoimmune inflammatory eye disease; allergy; ss.
OS	Homo sapiens.
XX	
PN	WO200009552-A1.
XX	
PD	24-FEB-2000.
XX	
PF	13-AUG-1999; 99WO-US18298.
XX	
PR	14-AUG-1998; 98US-0096622.
PR	17-AUG-1998; 98US-0096815.
PR	04-SEP-1998; 98US-0099229.
PR	23-OCT-1998; 98US-0105368.
PR	08-JAN-1999; 99US-0115234.
PR	12-FEB-1999; 99US-0119931.
PR	18-FEB-1999; 99US-0120575.
PR	30-APR-1999; 99US-0132020.
PR	11-AUG-1999; 99US-0096622.
PA	(GEMV ) GENETICS INST INC.
XX	
PI	Jacobs K, MCCoy JM, LaValle ER, Collins-Racie LA, Evans C;
PI	Menderg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
PI	Wong GG, Clark HF, Fechtel K;
XX	
DR	WP1: 2000-205979/18.
XX	P-PSDB; AAY94910.
PT	New polynucleotides encoding secreted proteins, which may have e.g.
PT	nutritional, chemokine, immune stimulating or suppressing,
PT	hematopoiesis regulating, tissue growth, activin/inhibin
PT	antiinflammatory or tumor inhibition activity -
XX	
PS	Claim 34; Page 493; 641pp; English.
XX	
CC	AAA16618 to AAA16697 encode the human secreted proteins given in
CC	AAY9498 to AAY94980, isolated from human adult brain, adult thyrdo,
CC	adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney
CC	adult placenta, adult testis, whole embryo, adult cartilage, kidney
CC	foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
CC	and adult bladder, cDNA libraries. The polynucleotides and proteins are
CC	predicted to have biological activities which would make them suitable
CC	for treating, preventing or ameliorating medical conditions in humans
CC	and animals. The polynucleotides can be used as markers for tissues in
CC	which the protein is preferentially expressed, as molecular weight
CC	markers on Southern gels, and as chromosome markers or tags to identify
CC	chromosomes or to map gene positions. The proteins can be used in the
CC	treatment of immune deficiencies and disorders, such as severe combined
CC	immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC	infections. These infections include human immunodeficiency virus (HIV),
CC	hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC	candidiasis. The proteins can be used to treat autoimmune disorders such
CC	as connective tissue disease, multiple sclerosis, systemic lupus
CC	erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC	



CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host-disease and CC autoimmune inflammatory eye disease. The proteins can also be used to CC treat allergic conditions, such as asthma. AA116568 to AA116774 represent CC probes for the human secreted proteins from the present invention.

Sequence 3370 BP; 1031 A; 537 C; 676 G; 1126 T; 0 other;

Query Match	2.1%;	Score 39.2;	DB 21;	Length 3370;
Best Local Similarity	47.9%;	Pred. No. 0.87;		
Matches 113;	Conservative 0;	Mismatches 123;	Indels 0;	Gaps 0;

QY	14	gaaaaggtgccatggagtggtgttataatglttcaaggtctccctacccaacatcgaata	73
Db	2571	GAATAAAGATTAATCACTTTAGACATTCACCTTTAAATAGTTTACTTAACTCAAAATG	2512
QY	74	ctctcgtagagaggtgcaatcgatcctatcaggacattccatctaacctcttggactt	133
Db	2511	TTTAATTAATAAACAGATATCTTCCACTTTAAACACTTTTGCTTTCTACTGTACAGTAATT	2452
QY	134	ttatgtgtatgcttcccaattctctctgtgtcctaattccctcgagtcgtatgctaatg	193
Db	2451	GCATTTGAGAGAGTACACCTTCCTGTTCAACCTGATCTCTTTGGATGGAAATTAAGATG	2392
QY	194	atcccttataataagtcgcaatctacacagcagcccttagcaacatatcgyaagc	249
Db	2391	TAACTGATAGTTTAAAGATTAATTAATGGAAGTTGGTGCACATCAATGATGACGC	2336

RESULT	9
AAK66361/c	
ID	AAK66361 standard; DNA; 38771 BP.

AA AAK66361;

DT 06-NOV-2001 (first entry)

DE	Human immune/haematopoietic antigen genomic sequence	SEQ ID NO:21173.
	<p>           1            2            3            4            5            6            7            8            9            10            11            12            13            14            15            16            17            18            19            20            21            22            23            24            25            26            27            28            29            30            31            32            33            34            35            36            37            38            39            40            41            42            43            44            45            46            47            48            49            50            51            52            53            54            55            56            57            58            59            60            61            62            63            64            65            66            67            68            69            70            71            72            73            74            75            76            77            78            79            80            81            82            83            84            85            86            87            88            89            90            91            92            93            94            95            96            97            98            99            100            101            102            103            104            105            106            107            108            109            110            111            112            113            114            115            116            117            118            119            120            121            122            123            124            125            126            127            128            129            130            131            132            133            134            135            136            137            138            139            140            141            142            143            144            145            146            147            148            149            150            151            152            153            154            155            156            157            158            159            160            161            162            163            164            165            166            167            168            169            170            171            172            173            174            175            176            177            178            179            180            181            182            183            184            185            186            187            188            189            190            191            192            193            194            195            196            197            198            199            200            201            202            203            204            205            206            207            208            209            210            211            212            213            214            215            216            217            218            219            220            221            222            223            224            225            226            227            228            229            230            231            232            233            234            235            236            237            238            239            240            241            242            243            244            245            246            247            248            249            250            251            252            253            254            255            256            257            258            259            260            261            262            263            264            265            266            267            268            269            270            271            272            273            274            275            276            277            278            279            280            281            282            283            284            285            286            287            288            289            290            291            292            293            294            295            296            297            298            299            300            301            302            303            304            305            306            307            308            309            310            311            312            313            314            315            316            317            318            319            320            321            322            323            324            325            326            327            328            329            330            331            332            333            334            335            336            337            338            339            340            341            342            343            344            345            346            347            348            349            350            351            352            353            354            355            356            357            358            359            360            361            362            363            364            365            366            367            368            369            370            371            372            373            374            375            376            377            378            379            380            381            382            383            384            385            386            387            388            389            390            391            392            393            394            395            396            397            398            399            400            401            402            403            404            405            406            407            408            409            410            411            412            413            414            415            416            417            418            419            420            421            422            423            424            425            426            427            428            429            430            431            432            433            434            435            436            437            438            439            440            441            442            443            444            445            446            447            448            449            450            451            452            453            454            455            456            457            458            459            460            461            462            463            464            465            466            467            468            469            470            471            472            473            474            475            476            477            478            479            480            481            482            483            484            485            486            487            488            489            490            491            492            493            494            495            496            497            498            499            500            501            502            503            504            505            506            507            508            509            510            511            512            513            514            515            516  </p>	

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer; cytostatic; gene therapy; vaccine; metastasis; ds.

..... Homo sapiens.

PN WO200157182-A2.

PD 09-AUG-20

PF 17-JAN-2001; 2001WO-US01354.

PR	31-JAN-2000;	2000US-017906.5.
PR	04-FEB-2000;	2000US-018062.8.
PR	24-FEB-2000;	2000US-018466.4.
PR	02-MAR-2000;	2000US-018635.0.
PR	16-MAR-2000;	2000US-018987.4.
PR	17-MAR-2000;	2000US-019007.6.
PR	18-APR-2000;	2000US-019812.3.
PR	19-MAY-2000;	2000US-02051.5.
PR	07-JUN-2000;	2000US-020946.7.
PR	28-JUN-2000;	2000US-021488.6.
PR	30-JUN-2000;	2000US-021513.5.
PR	07-JUL-2000;	2000US-021654.7.
PR	07-JUL-2000;	2000US-021688.0.
PR	11-JUL-2000;	2000US-021748.7.
PR	11-JUL-2000;	2000US-021749.6.
PR	14-JUL-2000;	2000US-021829.0.
PR	26-JUL-2000;	2000US-022096.3.
PR	26-JUL-2000;	2000US-022096.4.
PR	14-AUG-2000;	2000US-022451.8.
PR	14-AUG-2000;	2000US-022451.9.
PR	14-AUG-2000;	2000US-022521.3.
PR	14-AUG-2000;	2000US-022521.4.

PR	14-AUG-2000	2000US-0225366
PR	14-AUG-2000	2000US-0225567
PR	14-AUG-2000	2000US-0225568
PR	14-AUG-2000	2000US-0225570
PR	14-AUG-2000	2000US-0225571
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226679
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226682
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227709
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231342
PR	08-SEP-2000	2000US-0231343
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232081
PR	08-SEP-2000	2000US-0232080
PR	11-SEP-2000	2000US-0231868
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233325
PR	21-SEP-2000	2000US-0234377
PR	21-SEP-2000	2000US-0234374
PR	23-SEP-2000	2000US-0234998
PR	23-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236372
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239393
PR	13-OCT-2000	2000US-0239397
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-0241186
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241816
PR	01-NOV-2000	2000US-0244517
PR	08-NOV-2000	2000US-0246675
PR	08-NOV-2000	2000US-0246675
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524

Query Match	2.0%;	Score 38.2;	DB 22;	Length 38771;
Best Local Similarity	50.8%;	Pred. No. 7;		
Matches 91;	Conservative	0;	Mismatches 88;	Indels 0;
				Gaps 0;
QY 16	aaagagtcacgtatgtgttataatgtctcaagtcgtccctacctaacaattgaataact	75		
Db 2840	AAATGATATATATCTTTGTGGTTTGGATTTCGATTCCCTAATGACATNAAGATATTAAACAG	2781		
QY 76	tgtctagaggagatgtgaacacatgcataatggactatccataactctcttggacttt	135		
Db 2780	CTTTTCACACAGCTTATTTGGTCGCATGATATGCTCTTTTAGAGACAGTCGTATTTCATATCC	2721		
QY 136	attgtttagctctcccatcttctctgttgcctaattccctgtgtctgtgagtcataaga	194		
Db 2720	ATTTTGTAACTGGCTTATCTTTTATTGTGTGAATTCNAAGATTCCTTAGACATCTGA	2662		
RESULT 10				
AAK68883/C				
ID AAK68883	standard; DNA; 38771 BP.			
XX AC AAK68883;				
XX DT 06-NOV-2001	(first entry)			
DE Human	immune/haematopoietic antigen genomic sequence SEQ ID NO:23695.			
XX KW Human; immune; hematopoietic; immune/haematopoietic antigen; cancer;				
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.				
XX OS Homo sapiens.				
XX PN WO200157182-A2.				
XX PD 09-AUG-2001.				
XX PF 17-JAN-2001;	2001WO-US01354.			
XX PR 31-JAN-2000;	2000US-0179065.			
PR 04-FEB-2000;	2000US-0180628.			
PR 24-FEB-2000;	2000US-0184664.			
PR 02-MAR-2000;	2000US-0186350.			
PR 16-MAR-2000;	2000US-0189874.			
PR 17-MAR-2000;	2000US-0190076.			
PR 18-APR-2000;	2000US-0198123.			
PR 19-MAY-2000;	2000US-0205515.			
PR 07-JUN-2000;	2000US-0209467.			
PR 28-JUN-2000;	2000US-0214886.			
PR 30-JUN-2000;	2000US-0215135.			
PR 07-JUL-2000;	2000US-0216647.			
PR 07-JUL-2000;	2000US-0216880.			
PR 11-JUL-2000;	2000US-0217487.			
PR 14-JUL-2000;	2000US-0217496.			
PR 14-JUL-2000;	2000US-0218290.			
PR 26-JUL-2000;	2000US-0220963.			
PR 26-JUL-2000;	2000US-0220964.			
PR 14-AUG-2000;	2000US-0224518.			
PR 14-AUG-2000;	2000US-0224519.			
PR 14-AUG-2000;	2000US-0225213.			
PR 14-AUG-2000;	2000US-0225214.			
PR 14-AUG-2000;	2000US-0225266.			
PR 14-AUG-2000;	2000US-0225267.			
PR 14-AUG-2000;	2000US-0225268.			
PR 14-AUG-2000;	2000US-0225447.			
PR 14-AUG-2000;	2000US-0225757.			
PR 14-AUG-2000;	2000US-0225758.			
PR 14-AUG-2000;	2000US-0225759.			
PR 18-AUG-2000;	2000US-0226279.			
PR 22-AUG-2000;	2000US-0226681.			
PR 22-AUG-2000;	2000US-0226868.			





PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
XX Disclosure; SEQ ID NO 24416; 3071pp + Sequence listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 38771 BP; 9791 A; 8245 C; 8364 G; 10846 T; 1525 other;  
SO

Query Match 2.0%; Score 38.2; DB 22; Length 38771;  
Best Local Similarity 50.8%; Pred. No. 7;  
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 16 aaagagtcgatgagtggttataatgtaagtcctccatccatcaactgaact 75  
Db 35932 aaagaatcatctggttggttgcattcctcctaatagtaagaatataacag 35991

QY 76 tgcataagagagtcgaacatgatcatagagcatatcatcattcttgagcttt 135  
Db 35992 ctttcacagagctattgagcagatgatactctctttaaagaacgtctcatcatatcc 36051

QY 136 attgtgtagtcctccatctctctgtgctaaatccctgagtcgttaggtcataaga 194  
Db 36052 atttgaactgggtatattcttattgttgtaattctaagagctctttagacatcga 36110

RESULT 12  
AAK71499/G  
ID AAK71499 standard; DNA: 38771 BP.

XX  
AC AAK71499;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26311.  
XX  
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225457.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226682.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.

PR	14-SEP-2000	2000US-02324200
PR	14-SEP-2000	2000US-02324201
PR	14-SEP-2000	2000US-02330603
PR	14-SEP-2000	2000US-02330604
PR	14-SEP-2000	2000US-02330605
PR	21-SEP-2000	2000US-02342223
PR	21-SEP-2000	2000US-02342274
PR	25-SEP-2000	2000US-02349597
PR	25-SEP-2000	2000US-02354998
PR	26-SEP-2000	2000US-02354884
PR	27-SEP-2000	2000US-02356334
PR	27-SEP-2000	2000US-02356336
PR	29-SEP-2000	2000US-02363627
PR	29-SEP-2000	2000US-02363627
PR	29-SEP-2000	2000US-02363667
PR	29-SEP-2000	2000US-02363669
PR	29-SEP-2000	2000US-02363670
PR	02-OCT-2000	2000US-02370302
PR	02-OCT-2000	2000US-02370307
PR	02-OCT-2000	2000US-02370308
PR	02-OCT-2000	2000US-02370309
PR	02-OCT-2000	2000US-02370340
PR	13-OCT-2000	2000US-02370400
PR	13-OCT-2000	2000US-02393937
PR	13-OCT-2000	2000US-02409600
PR	20-OCT-2000	2000US-02412221
PR	20-OCT-2000	2000US-02417851
PR	20-OCT-2000	2000US-02417856
PR	20-OCT-2000	2000US-02417867
PR	20-OCT-2000	2000US-02417877
PR	20-OCT-2000	2000US-02418008
PR	20-OCT-2000	2000US-02418009
PR	20-OCT-2000	2000US-02418026
PR	01-NOV-2000	2000US-02446177
PR	01-NOV-2000	2000US-02446174
PR	08-NOV-2000	2000US-02464574
PR	08-NOV-2000	2000US-02464575
PR	08-NOV-2000	2000US-02464576
PR	08-NOV-2000	2000US-02464577
PR	08-NOV-2000	2000US-02464578
PR	08-NOV-2000	2000US-02465323
PR	08-NOV-2000	2000US-02465322
PR	08-NOV-2000	2000US-02465609
PR	08-NOV-2000	2000US-02465610
PR	08-NOV-2000	2000US-02466111
PR	17-NOV-2000	2000US-02462110
PR	17-NOV-2000	2000US-02462112
PR	17-NOV-2000	2000US-02462113
PR	17-NOV-2000	2000US-02462114
PR	17-NOV-2000	2000US-02462115
PR	17-NOV-2000	2000US-02462116
PR	17-NOV-2000	2000US-02462117
PR	17-NOV-2000	2000US-02462118
PR	17-NOV-2000	2000US-02462119
PR	17-NOV-2000	2000US-02462120
PR	17-NOV-2000	2000US-02462121
PR	17-NOV-2000	2000US-02462122
PR	17-NOV-2000	2000US-02462123
PR	17-NOV-2000	2000US-02462124
PR	17-NOV-2000	2000US-02462125
PR	17-NOV-2000	2000US-02462126
PR	17-NOV-2000	2000US-02462127
PR	17-NOV-2000	2000US-02462128
PR	17-NOV-2000	2000US-02462129
PR	17-NOV-2000	2000US-02462130
PR	17-NOV-2000	2000US-02462131
PR	17-NOV-2000	2000US-02462132
PR	17-NOV-2000	2000US-02462133
PR	17-NOV-2000	2000US-02462134
PR	17-NOV-2000	2000US-02462135
PR	17-NOV-2000	2000US-02462136
PR	17-NOV-2000	2000US-02462137
PR	17-NOV-2000	2000US-02462138
PR	17-NOV-2000	2000US-02462139
PR	17-NOV-2000	2000US-02462140
PR	17-NOV-2000	2000US-02462141
PR	17-NOV-2000	2000US-02462142
PR	17-NOV-2000	2000US-02462143
PR	17-NOV-2000	2000US-02462144
PR	17-NOV-2000	2000US-02462145
PR	17-NOV-2000	2000US-02462146
PR	17-NOV-2000	2000US-02462147
PR	17-NOV-2000	2000US-02462148
PR	17-NOV-2000	2000US-02462149
PR	17-NOV-2000	2000US-02462150
PR	17-NOV-2000	2000US-02462151
PR	17-NOV-2000	2000US-02462152
PR	17-NOV-2000	2000US-02462153
PR	17-NOV-2000	2000US-02462154
PR	17-NOV-2000	2000US-02462155
PR	17-NOV-2000	2000US-02462156
PR	17-NOV-2000	2000US-02462157
PR	17-NOV-2000	2000US-02462158
PR	17-NOV-2000	2000US-0246

PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-)	HUMAN GENOME SCI INC.
PI	Rosen CA,	Barash SC, Ruben SM;
PI	WPI;	2001-483426/52.
DR		
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis	
XX		
PS	Disclosure:	SEQ ID NO 26311; 3071bp + Sequence Listing; English.
XX		
CC	AAK64951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytosolic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins, and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/haematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169	
CC	represent sequences used in the exemplification of the present invention.	
XX		
SD	Sequence 38771 BP; 10846 A; 8364 C; 8245 G; 9791 T; 1525 other:	
	Query Match	2.0%; Score 38.2; DB 22; Length 38771;
	Best Local Similarity	50.8%; Pred. No. 7;
	Matches	91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY	16	aaaagtcacatgagatggtttataatggtcaaggctctccctcaacaacttgaact 75
DB	2840	AAATGATATATCATGTTGGTTTGCATTTGCCATTTCCCAATGAGCTAAGAATATTAACAG 2781
QY	76	tgtctagagagttgaacatcgatctatgagactatccatcaactctctttgacctt 135
DB	2780	CTTTTCACGAGCTATATTTGGCCATGATATATGCTCTTTTGAAGAACTCTCATTTATATGCC 2721
QY	136	attgttgatgctctccatctctctgtgtgccaattcccttagctgtagtgactaaga 194
DB	2720	ATTTGTCTAACTGGGTATCTTTTATTTGTTGAAATCTAAGAGTTCCTTTAGACATCTGA 2662
	RESULT 13	
ID	AAK72925	AAK72925
AC	AAK72925 standard; DNA: 38771 BP.	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27737.	
XX		
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KW	cytostatic; gene therapy; vaccine; metastasis; ds.	
XX		
OS	Homo sapiens	

XX WO200157182-A2.  
XX  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX





CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins, and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/haematopoietic-related diseases, especially
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC	to AAK87654 represent human immune/haematopoietic antigen genomic
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC	represent sequences used in the exemplification of the present invention.
XX	
SO	Sequence 38771 BP; 10846 A; 8364 C; 8245 G; 9791 T; 1525 other;
Query Match	
Best Local Similarity	2.0%; Score 38.2; DB 22; Length 38771;
Matches	91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY	16 aaaagtcacatgagatggtttataatgltcaaggtctccctacaacatgaaact 75
DB	2840 AAAATGATATATCATTTGTGTGATTTGCATTTCCCAATAGACTMAAGATATTAACAG 2781
QY	76 tgctagaggaattgacaactgcgatctatgaggaactatccatctaactccttggaacttt 135
DB	2780 CTTTTCACGACCTTATTTGGCCATGATATATGTCCTCTTTTGAAGAACTGCTATTCATATACC 2721
QY	136 attgtgtgattctcccaattctctgtgtgccaattccctgactgctgaggtcataaga 194
DB	2720 ATTGTGTAACATGGGCTTATCTTTTATTGTGTAATCTTAAGAGTCTTTAACAATCTGA 2662
RESULT 15	
AAK80325	
ID	AAK80325 standard; DNA; 38771 BP.
XX	
AC	AAK80325;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35137.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225256.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251968.  
 PR 05-DEC-2000; 2000US-0251969.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251858.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-483426/52.  
 Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and  
 metastasis -  
 Disclosure; SEQ ID NO 35137; 3071bp + Sequence Listing; English.  
 AA54951 to AA64702 encode the human immune/hematopoietic antigen (I)  
 amino acid sequences given in AA62170 to AA61921. (I) have cytostatic  
 activity, and can be used in gene therapy and vaccine production. (I)  
 proteins and polynucleotides may be used in the prevention, diagnosis and  
 treatment of diseases associated with inappropriate (I) expression. For  
 example, they may be used to treat disorders associated with decreased  
 expression by rectifying mutations or deletions in a patient's genome  
 that affect the activity of (I) by expressing inactive proteins or to  
 supplement the patient's own production of (I). Additionally, (I)  
 polynucleotides may be used to produce the secreted (I), by inserting  
 the nucleic acids into a host cell and culturing the cell to express the  
 protein. (I) proteins and polynucleotides may be used to prevent,

CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AA664703  
CC to AA876594 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AA554542 to AA554950 and  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 38771 BP; 9791 A; 8245 C; 8364 G; 10846 T; 1525 other;

Query Match	2.0%;	Score 38.2;	DB 22;	Length 38771;
Best Local Similarity	50.8%;	Pred. No. 7;		
Matches 91; Conservative	0;	Mismatches 88;	Indels 0;	Gaps 0;

[illegible]

Search completed: August 20, 2002, 01:14:04  
Job time: 4479 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:55:25 ; Search time 55.3 Seconds  
(without alignments)  
8470.580 Million cell updates/sec

Title: US-09-662-812-1  
Perfect score: 1907  
Sequence: 1 gtgctgatttgaaag.....ctatcaagagcgtgtaa 1907

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/Dackfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.8	1.9	4940	US-08-484-105-1	Sequence 1, Appli
2	36.8	1.9	4940	US-08-484-106-1	Sequence 1, Appli
3	36	1.9	7218	US-08-232-463-14	Sequence 14, Appli
4	34.8	1.8	1201	US-08-286-872-1	Sequence 1, Appli
5	34.4	1.8	4301	US-08-121-446-3	Sequence 3, Appli
6	33.4	1.8	4706	US-08-331-081B-5	Sequence 5, Appli
7	33.4	1.8	4744	US-08-331-081B-3	Sequence 3, Appli
8	33.2	1.7	14602	US-08-597-236-1	Sequence 1, Appli
9	33.2	1.7	14602	US-08-746-682A-1	Sequence 1, Appli
10	33	1.7	767	US-08-998-416-472	Sequence 472, App
11	32.6	1.7	3621	US-09-019-201A-1	Sequence 1, Appli
12	32.6	1.7	31571	US-08-323-443B-1	Sequence 1, Appli
13	32.6	1.7	53526	US-08-658-136-2	Sequence 2, Appli
14	32.6	1.7	53577	US-08-658-136-1	Sequence 1, Appli
15	32.2	1.7	9636	US-08-323-170B-1	Sequence 1, Appli
16	32.2	1.7	9636	US-08-954-441-1	Sequence 1, Appli
17	32	1.7	4376	US-08-119-125A-1	Sequence 1, Appli
18	32	1.7	4380	US-07-582-945-1	Sequence 1, Appli
19	32	1.7	4380	US-08-453-141-1	Sequence 1, Appli
20	32	1.7	4380	US-08-293-314-1	Sequence 1, Appli
21	32	1.7	6744	US-08-119-125A-2	Sequence 2, Appli
22	31.8	1.7	6136	US-08-137-252-1	Sequence 1, Appli
23	31.8	1.7	30001	US-08-125-468-1	Sequence 1, Appli
24	31.8	1.7	30001	US-08-474-933-1	Sequence 1, Appli
25	31.2	1.6	1519	US-08-850-910A-44	Sequence 44, Appli
26	30.6	1.6	1275	US-08-793-666-9	Sequence 9, Appli
27	30.6	1.6	7218	US-08-232-463-14	Sequence 14, Appli

28	30.6	1.6	7505	US-09-078-294-13	Sequence 13, Appli
29	30.6	1.6	8600	US-09-457-708-1	Sequence 1, Appli
30	30.4	1.6	762	US-08-998-416-462	Sequence 462, App
31	30.2	1.6	1209	US-08-822-774-18	Sequence 18, Appl
32	30.2	1.6	1209	US-09-632-711-18	Sequence 18, Appl
33	30.2	1.6	2738	US-08-795-868-17	Sequence 17, Appl
34	30.2	1.6	2738	US-09-303-069-17	Sequence 17, Appl
35	30.2	1.6	10395	US-08-245-809-3	Sequence 3, Appli
36	30.2	1.6	10798	US-08-107-748-2	Sequence 2, Appli
37	30.2	1.6	10798	PCT-US82-01385-2	Sequence 1025, Ap
38	30	1.6	1995	PCT-US96-05320A-1025	Sequence 1, Appli
39	30	1.6	3055	US-08-236-754-1	Sequence 1, Appli
40	30	1.6	8655	US-09-075-272-1	Sequence 77, Appl
41	29.8	1.6	389	US-09-227-357-77	Sequence 3, Appli
42	29.8	1.6	2088	US-09-351-414-3	Sequence 3, Appli
43	29.6	1.6	960	US-08-245-511-3	Sequence 3, Appli
44	29.6	1.6	960	US-08-600-993A-3	Sequence 3, Appli
45	29.6	1.6	1999	US-08-961-083-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-484-105-1  
; Sequence 1, Application US/08484105  
; Patent No. 5589341  
; GENERAL INFORMATION:  
; APPLICANT: STILLMAN, Bruce  
; APPLICANT: BELL, Stephen P  
; APPLICANT: KOBAYASHI, Ryuji  
; APPLICANT: RINE, Jasper  
; APPLICANT: FOSS, Margit  
; APPLICANT: MCNALLY, Francis J  
; APPLICANT: LAURENSEN, Patricia  
; APPLICANT: HERSKOWITZ, Ira  
; APPLICANT: LI, Joachim J  
; APPLICANT: GAVIN, Kimberly  
; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/484,105  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard Aron  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 494-8700  
; TELEFAX: (415) 494-8771  
; TELEEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4940 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-484-105-1

1509 ttataaatctcaccacaagaatcttcgaaaaaaaacgcttcctcatgattatgcacaaaaagc 1568  
 ||||| || || ||||| ||||| ||||| || |||||  
 142 TTATTAATTTCCCTAAGATGACACAAAAAATGTCTCCAAAAAATTACCAAGAAAAA 201

1341 xx 1400  
1488 ggcgcatafcagacatctgtattataatcaccacaagaatccgaaaaaaaaacgtctc 1547

Db 1401 yyyyyyyyyyyyyyyyyyyyyyyGTACCAAATCTCTATCTCTTAAC 1460  
OY 1548 tccatgatatgcaaaaagccttaattcagaacctgttatcttgagacaggaaccc 1607  
Db 1461 tACTTCATAGATAGTAGTAATTACAGTGAGCTACATAGCCGTTTTTGAACGTAAGA 1520  
OY 1608 aaagaccatccagacatgcctatctac 1637  
Db 1521 TGGGTCTAGAGGAGTACGTACGCTACTACAC 1550

## RESULT 4

US-08-286-872-1/C  
; Sequence 1, Application US/08286872  
; Patent No. 5675061  
; GENERAL INFORMATION:  
; APPLICANT: Powers, Dennis A.  
; TITLE OF INVENTION: Isolation and Characterization  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: P.O. Box 60850  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-0850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,872  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/192,272  
; FILING DATE: 04-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 8600-0144.30  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1201 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: Coho salmon Growth Hormone Gene  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 65..697  
; OTHER INFORMATION: /note= "Growth Hormone Precursor"  
; FEATURE:  
; NAME/KEY: sig\_peptide  
; LOCATION: 65..130  
; OTHER INFORMATION: /note= "growth hormone signal"  
; OTHER INFORMATION: peptide"  
; FEATURE:  
; NAME/KEY: mat\_peptide  
; LOCATION: 131..697  
; OTHER INFORMATION: /note= "Growth Hormone"  
US-08-286-872-1

Query Match 1.8%; Score 34.8; DB 1; Length 1201;

Best Local Similarity 57.3%; Pred. No. 0.67;  
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
OY 1232 gtyacaatcatagatccctcaagaatcccagagagaggaagtgtctctcacaccga 1291  
Db 217 GTCAAAGTCATTGACATTTCTGAGCCATAGTGAGATGTTGACCCGACTGACCGC 158  
OY 1292 gftctgaagagagacatcccttaagcttctcttcgccttaagaagc 1341  
Db 157 GATGTTGAGAGGCCGTTGTTTCTATGCTGCCCTTGACTGACGGAAC 108

## RESULT 5

US-08-121-446-3  
; Sequence 3, Application US/08121446  
; Patent No. 6313276  
; GENERAL INFORMATION:  
; APPLICANT: IMURA, HIROO  
; APPLICANT: NAKAO, KAZUMA  
; APPLICANT: NAKANISHI, SHIGETADA  
; TITLE OF INVENTION: A HUMAN ENDOTHELIN RECEPTOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/121,446  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/911,684  
; FILING DATE: 10-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CIOTTI, THOMAS E.  
; REGISTRATION NUMBER: 21,013  
; REFERENCE/DOCKET NUMBER: 29900-20324.00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4301 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 238..1566  
US-08-121-446-3

Query Match 1.8%; Score 34.4; DB 4; Length 4301;  
Best Local Similarity 50.0%; Pred. No. 1.9;  
Matches 86; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

OY 797 tctctcaagatcaagaagctatctttagctttaggaagcttaagatggtcagagc 856  
Db 2431 TTTTTCAGGATATATTAATCTCTTTTTCATATCGTAGCTTAACCTGTTGTT 2490  
OY 857 tactacaataaaaaagaacatgcagagaccctgagtgtgagtcaacttaagcagagc 916  
Db 2491 TTGTCACTGTGTAATACTTACCTACATACACGCACTGATGATGATTAAGAGGAGGCGC 2550





```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/597,236
FILING DATE:
CLASSIFICATION: 426
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95201669.9
FILING DATE: 20-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fanucci A., Allan
REGISTRATION NUMBER: 30256
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14602 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 352..1803
OTHER INFORMATION: /product= "epsa"
FEATURE:
NAME/KEY: CDS
LOCATION: 1807..2535
OTHER INFORMATION: /product= "epsb"
FEATURE:
NAME/KEY: CDS
LOCATION: 2547..3239
OTHER INFORMATION: /product= "epsc"
FEATURE:
NAME/KEY: CDS
LOCATION: 3249..3995
OTHER INFORMATION: /product= "epsd"
FEATURE:
NAME/KEY: CDS
LOCATION: 4051..4731
OTHER INFORMATION: /product= "epse"
FEATURE:
NAME/KEY: CDS
LOCATION: 4898..5854
OTHER INFORMATION: /product= "epsf"
FEATURE:
NAME/KEY: CDS
LOCATION: 5854..6425
OTHER INFORMATION: /product= "epsg"
FEATURE:
NAME/KEY: CDS
LOCATION: 6425..7540
OTHER INFORMATION: /product= "epsh"
FEATURE:
NAME/KEY: CDS
LOCATION: 7736..8212
OTHER INFORMATION: /product= "epsi"
FEATURE:
NAME/KEY: CDS
LOCATION: 8221..9192
OTHER INFORMATION: /product= "epsj"
FEATURE:
NAME/KEY: CDS
LOCATION: 9285..10364
OTHER INFORMATION: /product= "epsk"
FEATURE:
NAME/KEY: CDS
LOCATION: 10392..11339
OTHER INFORMATION: /product= "epsl"
FEATURE:
NAME/KEY: misc.feature
LOCATION: 11302..12222
OTHER INFORMATION: (eps k) on nucleotides 10392-11339"
FEATURE:
NAME/KEY: CDS
LOCATION: 12233..13651
OTHER INFORMATION: /product= "epsm"
FEATURE:
NAME/KEY: misc.feature

```

```

LOCATION: 13732..14305
OTHER INFORMATION: //function= "CDS on the
OTHER INFORMATION: complementary strand"
OTHER INFORMATION: /product= "orf2"
FEATURE:
NAME/KEY: terminator
LOCATION: 230..252
FEATURE:
NAME/KEY: promoter
LOCATION: 274..302
FEATURE:
NAME/KEY: RBS
LOCATION: 340..345
US-08-597-236-1

Query Match      1.7%; score 33.2; DB 1; length 14602;
Best local similarity 57.8%; Pred. No. 9.2; Mismatches 43; Indels 0; Gaps 0
Matches 59; Conservative 0;

QY 1646 gtccacccagaagaacgctagcaagctcgttctgtgatattctagagacactagccacctcg 1705
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 6131 GACACACTGCAAGCAGCGACTCCAAATTACCTGGTGCAATATCGTAGTAACGGTAGTACCTAG 6072

QY 1706 aagctctccgaagaatattccgtttattgtatatacaactcgatgagc 1747
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 6071 ACGTCCACCGAATAAGACATTCTCTTTGTCGCGCAAGACG 6030

RESULT 9
US-08-746-682A-1/C
; Sequence 1, Application US/08746682A
; Patent No. 5786184
; GENERAL INFORMATION:
; APPLICANT: STINGELE, Francesca
; APPLICANT: MOLLET, Beat
; TITLE OF INVENTION: LACTIC BACTERIA PRODUCING
; TITLE OF INVENTION: EXOPOLYSACCHARIDES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americans
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746, 682A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/597, 236
; FILING DATE: 20-JUN-1995
; APPLICATION NUMBER: EP 95201669.9
; FILING DATE: 20-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Panucci A., Allan
; REGISTRATION NUMBER: 30256
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 352..1803

```

OTHER INFORMATION: /product= "epsa"  
FEATURE: NAME/KEY: CDS  
LOCATION: 1807..2535  
OTHER INFORMATION: /product= "epsb"  
FEATURE: NAME/KEY: CDS  
LOCATION: 2547..3239  
OTHER INFORMATION: /product= "epsc"  
FEATURE: NAME/KEY: CDS  
LOCATION: 3249..3995  
OTHER INFORMATION: /product= "epsd"  
FEATURE: NAME/KEY: CDS  
LOCATION: 4051..4731  
OTHER INFORMATION: /product= "epse"  
FEATURE: NAME/KEY: CDS  
LOCATION: 4898..5854  
OTHER INFORMATION: /product= "epsf"  
FEATURE: NAME/KEY: CDS  
LOCATION: 6425..7540  
OTHER INFORMATION: /product= "epsG"  
FEATURE: NAME/KEY: CDS  
LOCATION: 7736..8212  
OTHER INFORMATION: /product= "epsH"  
FEATURE: NAME/KEY: CDS  
LOCATION: 8221..9192  
OTHER INFORMATION: /product= "epsi"  
FEATURE: NAME/KEY: CDS  
LOCATION: 9285..10364  
OTHER INFORMATION: /product= "epsJ"  
FEATURE: NAME/KEY: CDS  
LOCATION: 10392..11339  
OTHER INFORMATION: /product= "epsK"  
FEATURE: NAME/KEY: misc.feature  
LOCATION: 11302..12222  
OTHER INFORMATION: /product= "CDS (eps I) covering CDS  
OTHER INFORMATION: (eps k) on nucleotides 10392-11339"  
FEATURE: NAME/KEY: CDS  
LOCATION: 12233..13651  
OTHER INFORMATION: /product= "epsM"  
FEATURE: NAME/KEY: misc.feature  
LOCATION: 13732..14505  
OTHER INFORMATION: /function= "CDS on the  
OTHER INFORMATION: complementary strand"  
OTHER INFORMATION: /product= "orfz"  
FEATURE: NAME/KEY: terminator  
LOCATION: 230..252  
FEATURE: NAME/KEY: promoter  
LOCATION: 274..302  
FEATURE: NAME/KEY: RBS  
LOCATION: 340..345  
US-08-746-682A-1

Query Match 1.7%; Score 33.2; DB 1; Length 14602;  
Best Local Similarity 57.8%; Pred. No. 9.2;  
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 1646 gtcccccagaagcgtagaagctatcgttgatattcgaagacactgccaactcg 1705

Db 6131 GACACACTGCAGAGCGCTCCATTACTGTGATATCGTAGACGTAGCTAG 6072  
QY 1706 aagttccgaagatccgtttatgtacacgtagacg 1747  
Db 6071 ACgtccaccgaaatgacattctctgttgcctgcgcagacg 6030

RESULT 10  
US-08-998-416-472  
Sequence 472, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippesen, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Reibischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIL  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwalls Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 472:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 767 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1337RP  
US-08-998-416-472

Query Match 1.7%; Score 33; DB 4; Length 767;  
Best Local Similarity 55.8%; Pred. No. 1.9;  
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 92 catgacatagggacatccatcactctcttgagacttattgtgtagcttccc 151  
Db 392 CATTGATATATATATATATTAATGTAAGTCAATTAATTTATTTATTTAGTCTAGTA 451  
QY 152 attctctgttgctaaatccctgagctgtgagtcataaagatccattat 204  
Db 452 ATATTTCATTATTAATAGTCTCCCTTTAATGTGATATTAATCTACCTAATAATAT 504

```

APPLICANT: LANDES, GREGORY M.
APPLICANT: BURN, TIMOTHY C.
APPLICANT: CONNORS, TIMOTHY D.
APPLICANT: DACKOWSKI, WILLIAM R.
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,443B
FILING DATE: 12-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, S. Peter
REGISTRATION NUMBER: 25,351
REFERENCE//DOCKET NUMBER: 0372/OA462
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 31571 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: PKD1 GENOMIC
US-08-323-443B-1

Query Match 1.7%; Score 32.6; DB 1; Length 31571;
Best Local Similarity 63.3%; Pred. No. 22;
Matches 50; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 351 gcacatcatcattgagcgagcctgagcgtatctcagaagccttgcagctgctcccaag 410
||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 5078 gcacatcattatcaccgaagcagcgtgtgacctcagaagcccgctctccacacgtgg 5019

QY 411 ctatgaaactgcagacc 429
|||| || || || || ||

Db 5018 AACTGACACACAGCTGCTCC 5000

RESULT 13
US-08-658-136-2/c
; Sequence 2, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/658,136
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GEN4-17.8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 53526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-658-136--2

```

```

1 APPLICATION NUMBER: US/08/658,136
2 FILING DATE:
3 CLASSIFICATION: 435
4 ATTORNEY/AGENT INFORMATION:
5 NAME: LASSEN, ELIZABETH
6 REGISTRATION NUMBER: 31,845
7 REFERENCE/DOCKET NUMBER: GENA-17.8
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 508-872-8400
10 TELEFAX: 508-872-5415
11 INFORMATION FOR SEQ ID NO: 1:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 53577 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: single
16 TOPOLOGY: linear
17 MOLECULE TYPE: DNA (genomic)
18
19 US-08-658-136-1

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
US-08-323-170B-1

```

```

Query Match          1.7%; Score 32.2; DB 1; Length 9636;
Best Local Similarity 49.7%; Pred. No. 15;
Matches 82; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

Qy 16 aaaggtccatgcatgtgttaataatgtcaaggctccccaacaaactgaatct 75
   |||||
Db 6967 AATATGTTCACTACTTCTTATATATCTTATATATTAATAAATTAATAACA 6908
   |||||

Qy 76 tgcctagagaggtlgaacatcgatcgaactlgaactlgaactlgaactl 135
   |||||
Db 6907 TCGTTGTGGAATATAGCAAAACCTTTAGGACATTTCAATCCAAATATCTCTTTT 6848
   |||||

Qy 136 atgtgtagctctccatctctctctgtgtgtaaatlccctgaagtc 180
   |||||
Db 6847 AATTTTAACAGTGCACCTTTTCTTCTTACTTGACTATTTAATTC 6803
   |||||

```

Search completed: August 20, 2002, 01:09:12  
Job time: 4427 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 19, 2002, 23:00:25 ; Search time 1586.97 Seconds  
(without alignments)  
16218.765 Million cell updates/sec

Title: US-09-662-812-1  
Perfect score: 1907  
Sequence: 1 gtgctgatttgaaag.....ctcatcaagagcgtgaa 1907

Scoring table: IDENTITY NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estlum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.8	2.5	1101	12	CNS0039G
2	42.2	2.2	1101	12	CNS017KX
3	39.6	2.1	454	9	AI742802
4	39.6	2.1	564	9	AW216328
5	39.6	2.1	579	12	AO512496
6	39.6	2.1	719	12	AZ058799
7	39.2	2.1	364	10	BF943308
8	39.2	2.1	411	10	BI343252
9	39.2	2.1	452	9	AW952059
10	39.2	2.1	454	9	AI689130
11	39.2	2.1	455	9	AW999085
12	39.2	2.1	532	9	AI684725
13	39.2	2.1	733	10	BE869217
14	38.6	2.0	339	12	AZ911277
15	38.6	2.0	487	12	AZ074909
16	38.4	2.0	466	10	BE927675
17	38.4	2.0	618	10	BF305035

C 18	38.4	2.0	724	10	BG479306	BG479306 602526335
C 19	38.2	2.0	509	12	AO695687	AO695687 HS_2159_A
C 20	38.2	2.0	685	9	AV404653	AV404653 AV404653
C 21	38.2	2.0	725	12	BH467155	BH467155 BOROK53TR
C 22	38.2	2.0	728	12	BH466366	BH466366 BORON53TR
C 23	38.2	2.0	1101	12	CNS016B8	AL106526 Drosophila
C 24	38.2	2.0	807	12	AZ528490	AZ528490 ENTNC64TR
C 25	38	2.0	845	12	BH139541	BH139541 ENTNC88TR
C 26	37.8	2.0	385	12	AO142345	AO142345 HS_3064_A
C 27	37.8	2.0	464	9	AV707816	AV707816 AV707816
C 28	37.6	2.0	898	9	AL574540	AL574540 AL574540
C 29	37.6	2.0	1046	12	CNS0407V	AL307444 Tetradon
C 30	37.6	2.0	513	12	AO718260	AO718260 HS_5513_B
C 31	37.2	2.0	207	9	AI039088	AI039088 ox27501.s
C 32	37.2	2.0	275	10	D61175	D61175 HUM164D03B
C 33	37.2	2.0	647	12	BH577763	BH577763 BOGA877TR
C 34	37.2	2.0	725	12	AZ200838	AZ200838 SP_1011_A
C 35	37	1.9	533	12	AQ420915	AQ420915 RPCI-11-2
C 36	37	1.9	851	10	BF132845	BF132845 601645986
C 37	36.8	1.9	527	9	AJ398606	AJ398606 AJ398606
C 38	36.8	1.9	730	12	AZ519102	AZ519102 RPCI-11-7
C 39	36.8	1.9	900	9	AU067211	AU067211 AU067211
C 40	36.6	1.9	479	12	AQ002804	AQ002804 CIT-HSP-2
C 41	36.6	1.9	542	10	BI935367	BI935367 EST55256
C 42	36.6	1.9	740	10	BI926005	BI926005 EST545894
C 43	36.6	1.9	794	10	BI935368	BI935368 EST55257
C 44	36.4	1.9	421	9	AI267708	AI267708 ag93109.x
C 45	36.4	1.9	461	9	AW365573	AW365573 MR0-HR007

## ALIGNMENTS

RESULT 1  
CNS0039G 1101 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL063921  
VERSION AL063921.1 GI:4941778  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;  
Mecoptera; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
AUTHORS 1 (bases 1 to 1101)  
TITLE Genoscope.  
JOURNAL Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammosier in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source  
1. .1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="RPCI-98"







Gaps: 0;

ca 93  
||  
CA 270  
at 153

210

222-JAN-2001  
sequence.  
eostomi;

ones, M.R.,  
F.F.,  
ppson, D.H.,  
O'Hare  
and  
pressed

Paulo-SP,  
er Genome  
M4 - NN0091 -

for  
into"

001  
e.

me  
91 -



QY	194	accccttatataagcgaatctacacagcgccttagcacatatctggaagc	249
Db	272	GCATTGTAGAGGTACACTTCTGTCTTCCAAACTGTATCTTCTTTCGAGGAAATTAAGATG	331
QY	194	accccttatataagcgaatctacacagcgccttagcacatatctggaagc	249
Db	332	TAACTGTATAGCTTTTAAAGATAATTAATGGAAGTGTGTCACACTAAGATGACAGC	387
RESULT	11		
LOCUS	AM999085	455 bp	mrna linear EST 05-JUN-2000
DEFINITION	MR0-BN0070-010300-003-e11 BN0070	Homo sapiens	cdna, mRNA sequence.
ACCESSION	AM999085		
VERSION	AM999085.1	GI:8259319	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldstein, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brustein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R., Reis, L.F., de Souza, S.J. and Simpson, A.J.		
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7), 3491-3496	(2000)
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be found in the following URL (http://www.ludwig.org.br/scripts/gethtml.pl?l=613-MR0-BN0070-010300-003-e11&l3=2000-03-01&l4=1) Seq primer: puc 18 forward High quality sequence start: 13 High quality sequence stop: 455. Location/Qualifiers 1..455 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="BN0070" /dev_stage="Adult" /note="Organ: breast; normal; Vector: puc18; site:1: Smat: site:2: Smat: A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."		
BASE COUNT	163 a	75 c	64 g 153 t
ORIGIN			
Query Match	2.1%	Score 39.2;	DB 9; Length 455;
Best Local Similarity	47.9%	Pred. No. 5.9;	
Matches 113;	Conservative 0;	Mismatches 123;	Indels 0; Gaps 0;
QY	14	gaaaaagtcacatgagtggtttatataatgttcaagcgtcccatccaacatgaata	73
Db	192	GAAATTAAGATTAACACTTTTGAACATTCAGTTAAATGTAAGTACTTATTAATGTCATAAGT	251
QY	74	cttcctagaggaagtcagacatgcatactatggagcatcatcatcaatccttgcactt	133

Db	252	TTTATATAAAACAGCATCTTCGCATTTTAACACTTTCCTTTCAACTGTCACAGTAANTT	311
Oy	134	ttattgtgtagctcttcccatcttctctgttgcctaattccctagctgttagtcaag	193
Dc	312	GCATTGTAGAGACATACACTTCTGTCTTCAAACTGATCTTTTGGATGGAAATTAAGATG	371
Oy	194	atccctataatagtaagcaactacacacagcagcgcccttagcacacatctcgaagc	249
Dc	372	TAACTGTATACTTTTAAGATATAATTAATGAGGAGTGTGTCACACTAAGATGACAGC	427
RESULT	12		
AI684725			
DEFINITION	AI684725	532 bp	mRNA linear EST 07-MAR-2000
ACCESSION	W855d09.x1 Soares.NFL.T.GRC.S1		Homo sapiens cDNA clone
VERSION	IMAGE:2302961.3		mRNA sequence.
KEYWORDS	AI684725		
SOURCE	AI684725.1	GI:4896019	
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	1 (bases 1 to 532)		
COMMENT	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgaps-r@mail.nih.gov">cgaps-r@mail.nih.gov</a>		
	This clone is available royalty-free through LNL; contact the		
	IMAGE Consortium ( <a href="mailto:info@imgl.nl">info@imgl.nl</a> ) for further information.		
	Insert Length: 1160 Std Error: 0.00		
	Seq primer: -400p from Glibco		
	High quality sequence stop: 453.		

Query Match	Best Local Similarity	Score	DB	Length
Matches 113; Conservative 0; Mismatches 123; Indels 0; Gaps 0;	2.1%; 47.9%; Pred. No. 6.2;	39.2;	9;	532;
QY 14	gaaaaaggtccatcgagatgtgtttataatggtcaaggtctccctatccaacaacttgaata	73		
Db 147	GAAAAAAGATATATCTCTTTAGACATTCAGTTAAATGTAGTTATCTAAATCTCAAAAAG	206		
QY 74	cttgctagaaggaagttgcaacatcgacatactatgggaactlccatcaactctcttgaact	133		
Db 207	TTTATATAAAAACAAGATATCTTCTCCATTAAACACTTGGCTTCTTAAGCTTAACAGTAAT	266		
QY 134	ttaattgttagtcttcccatcttctcttctgttgcctaattccctcctgagctgttagtcataag	193		
Db 267	GCATTGTAGAGAGTACACTTCTGTCTCTCAAACTGTATCTTCTTTGGATGGAATTAAGATG	326		

Ox		194	atccctttataagtagcgaactcatcacagaaggacctgacaaatatcttgaagc	249 
Ddb	327	TACTGTATAGTTTAAGATAAATTAATGAATTGGTGCTCCACCTAACATGCACGC	382	
RESULT	13			
Locus	BEB69217/c			
DEFINITION	BEB69217	733 bp mRNA linear EST Z0-OCT-2000		
Accession	G01445402ZP1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3849388 5'			
Version	BEB69217	mRNA sequence.		
Keywords	BEB69217.1 GI:10317993			
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 733)			
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaabbs@email.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. GDMA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: CLAM956 row: m column: 05 High quality sequence stop: 598.			

BASE COUNT	237 a	110 c	136 g	249 t	1 others
ORIGIN					
Query Match	2.1%	Score 39.2;	DB 10;	Length 733;	
Best Local Similarity	47.9%;	Pred. No. 7;			
Matches 113;	Conservative 0;	Mismatches 123;	Indels 0;	Gaps 0;	
Qy	14	gaanaagtcctcagatgctgttataatgctcaagtcctccctccatccaacatgaaata	73		
Db	257	GAANAAGATATTCACCTTTAGACATTCAGTTAAATGTAGTTATCTTAATCTCAAAATG	198		
Qy	74	cttgctagagaagttggaacatcgacatctatggagactatccatctaactctcttggactt	133		
Db	197	TTTATTAATAAACAAGATATCTTCTCCATTTAAACACTTGGTTCTTCAACGTGTACAGTAATTT	138		
Qy	134	ttattgtgtagtctctccacttctcttggctgtaaatccctgagctgtaggtcataag	193		
Db	137	GCATGTGAGAGAGTACACTTCTGCTTCAAACTGTATCTTCTTGGATGGAAATTTAAGATG	78		
Qy	194	atcccttatataagtaagcaatctacacagcagcgcccttagcaacatatcttgyaaagc	249		
Db	77	TAACTGTATAGTTTAAAGATAAATAAATGGGAAGTTGGTCCAACTPAGATGACAGC	22		
RESULT 14					
AZ911277/c					
LOCUS	AZ911277	339 bp	DNA	linear	SSS 05-MAR-2001

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 487)	Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatsman, S., Aliret, B., Levins, M., McGinn, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C. M.	Mouse BAC End Sequences from Library RPCT-23	Unpublished (1999)	Other GSSS: RPCT-23-438022.TV
	Contact: Shaying Zhao			
	Department of Eukaryotic Genomics			
	The Institute for Genomic Research			
	9712 Medical Center Dr., Rockville, MD 20850, USA			
	Tel: 301 838 0200			
	Fax: 301 838 0208			
	Email: szhao@tigr.org			
	Clones are derived from the mouse BAC library RPCT-23. For BAC library availability, please contact Pieter de Jong			
	(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ( <a href="http://bacpac.med.buffalo.edu/orderingframe.htm">http://bacpac.med.buffalo.edu/orderingframe.htm</a> ) or from Resea ch Genetics (Info@resgen.com). BAC end page: <a href="http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html">http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html</a>			
	Plate: 438 row: 0 column: 22			
	Seq primer: SP6			
	Class: BAC ends.			
FEATURES	Location/Qualifiers			
SOURCE	1..487			
	/organism="Mus musculus"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10090"			
	/clone="RPCT-23-438022"			
	/clone_lib="RPCT-23"			
	/sex="Female"			
	/lab_host="DH10B"			
	/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site:1: EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)." " "			
BASE COUNT	105 a 100 c 140 g 142 t			
ORIGIN				
Query Match	2.0%; Score 38.6; DB 12; Length 487;			
Best Local Similarity	58.1%; Pred. No. 9;			
Matches	68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;			
OY	649 acgccttgagagctgaagatcttgatgcctatattctcgagatctcttaagctccaagaagaag 708			
Db	352 ACACGGAAGACTGGAAGAACTGCTCTCGAATATCAGGAGAAACGACACACAGCAACAG 293			
OY	709 cgcgattcgagagctgccacagctttgcagatcgagagatcaccaacaagaagccttct 765			
Db	292 GGCAGATGAGCTGCCAGACCAATGTGAAGGTAAAGCAAGCCCAAAACCAAGCTCTCT 236			

Search completed: August 20, 2002, 00:26:11  
Job time: 5146 sec

```
Search completed: August 20, 2002, 00:26:11
Job time: 5146 sec
```



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:22:21 ; Search time 19.49 Seconds

(Without alignments)  
2800.346 Million cell updates/sec

Title: US-09-662-812-2

Sequence: 1 MGFLHLTFLGLLCSLPISL.....TEGDAKNPVLAGLLIKIVE 568

Scoring table: OLIGO  
Gapop 60.0 , Gapept 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : PIR-71:\*\*

1: pirl:\*\*  
2: pirl:\*\*  
3: pirl:\*\*  
4: pirl:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	100.0	568	2 C72129	probable outer mem
2	568	100.0	568	2 G86493	probable leader pe
3	12	2.1	566	2 G71525	probable outer mem
4	10	1.8	566	2 A81682	conserved hypochet
5	9	1.6	408	1 E42409	biphenyl dioxigena
6	9	1.6	408	1 F41858	biphenyl dioxigena
7	9	1.6	445	2 T03578	probable glucose-i
8	9	1.6	502	2 T05135	hypothetical prote
9	8	1.4	134	2 S77785	hypothetical prote
10	8	1.4	137	2 C81695	cyclic nucleotide-
11	8	1.4	197	2 E95203	conserved hypochet
12	8	1.4	197	2 E98070	conserved hypochet
13	8	1.4	308	2 A12601	conserved hypochet
14	8	1.4	328	2 A97384	hypothetical prote
15	8	1.4	357	2 T03557	ribose transport s
16	8	1.4	414	2 S52618	hypothetical prote
17	8	1.4	416	2 G71096	hypothetical prote
18	8	1.4	434	2 A81597	similar glucose in
19	8	1.4	434	2 AD1234	glucose-inhibited
20	8	1.4	435	2 A72339	glucose-inhibited
21	8	1.4	435	2 A69632	glucose-inhibited
22	8	1.4	435	2 B83958	glucose-inhibited
23	8	1.4	435	2 F89898	glucose-inhibited
24	8	1.4	444	2 D97977	glucose-inhibited
25	8	1.4	447	2 A64934	Succinylarginine d
26	8	1.4	447	2 C90935	hypothetical prote
27	8	1.4	447	2 G85783	hypothetical prote
28	8	1.4	447	2 G86778	glucose inhibited
29	8	1.4	456	2 A95109	Gid protein [impor

30	8	1.4	483	2 T21327	hypothetical prote
31	8	1.4	566	2 T45162	succinate dehydrog
32	8	1.4	587	2 C86744	myosin-crossreacti
33	8	1.4	588	2 S66625	zeta-carotene desa
34	8	1.4	629	2 D96533	ARP protein [impor
35	8	1.4	629	2 S57614	ARP protein - Arab
36	8	1.4	666	2 S71032	flagellar hook-ass
37	8	1.4	667	2 F70682	probable membrane
38	8	1.4	671	2 A10816	conserved hypochet
39	8	1.4	761	2 H84950	ribonucleoside-dip
40	8	1.4	998	2 T32787	hypothetical prote
41	7	1.2	54	2 S60852	M protein precuso
42	7	1.2	76	2 E58933	ATP synthase prote
43	7	1.2	80	4 S60901	hypothetical prote
44	7	1.2	103	2 D72521	hypothetical prote
45	7	1.2	104	2 D72657	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

probable outer membrane leader peptide (comp) CPn0021 - Chlamydia pneumoniae (str C72129)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: C72129, B81542  
R:Kaldan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999  
M:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: C72129  
A:Molecule type: DNA  
A:Residues: 1-568 <ARN>  
A:Cross-references: GB:AE001587; GB:AE001363; NID:94376271; PIDN:AD18174.1; PID:943  
A:Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salz Nucleic Acids Res. 28, 1397-1406, 2000  
M:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR3  
A:Reference number: A81500; MUID:20150255  
A:Accession: B81542  
A:Molecule type: DNA  
A:Residues: 1-568 <REA>  
A:Cross-references: GB:AE002234; GB:AE002161; NID:97189667; PIDN:AR38558.1; PID:971  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CPn0021; CP0755

Query Match	100.0%;	Score 568;	DB 2;	Length 568;
Best Local Similarity	100.0%;	Pred. NO. 0;		
Matches 568;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MGFLHLTFLGLLCSLPISLVAKPESYGHITLITSTOSTOQALATYLEADAYGDHDF	60	
DB	1	MGFLHLTFLGLLCSLPISLVAKPESYGHITLITSTOSTOQALATYLEADAYGDHDF	60	
QY	61	VLRKIGEDYLKQSHSSDPQTRKSTIGAGLAGSSEALDVLQAMETADPLQQLLVLSAV	120	
DB	61	VLRKIGEDYLKQSHSSDPQTRKSTIGAGLAGSSEALDVLQAMETADPLQQLLVLSAV	120	
QY	121	SGHIGKTSDDLFPKALASPVVIREAVYRLANKTKFVIDHLHSFHKLPDEIOCLSA	180	
DB	121	SGHIGKTSDDLFPKALASPVVIREAVYRLANKTKFVIDHLHSFHKLPDEIOCLSA	180	
QY	181	IFRLTEESDAYIRDLAAKSAIRSATQIGEQKREPLTLRNLTLSASPOQDEAI	240	
DB	181	IFRLTEESDAYIRDLAAKSAIRSATQIGEQKREPLTLRNLTLSASPOQDEAI	240	
QY	241	LYALGKLKDGOSYNIKKOLQKPDVDTLAAQALIALGKEDALPVTKKALERRPAL	300	
DB	241	LYALGKLKDGOSYNIKKOLQKPDVDTLAAQALIALGKEDALPVTKKALERRPAL	300	

QY 301 YALRLPEIGIPIALPIFLTKNSEAKLNVALLEIGCDTPKLEXYTERLYOPHYNE 360  
|  
Db 301 YALRLPEIGIPIALPIFLTKNSEAKLNVALLEIGCDTPKLEXYTERLYOPHYNE 360  
QY 361 TLASFSGRGLQNMKRVNIIVPODERERLSTTRGLEQILFELRLPREALPCIT 420  
|  
Db 361 TLASFSGRGLQNMKRVNIIVPODERERLSTTRGLEQILFELRLPREALPCIT 420  
QY 421 KLASQKQTALTTAITSFSLTSHOQALDLFQAALPGEPIIRAYADAIYNLTKDPEKK 480  
|  
Db 421 KLASQKQTALTTAITSFSLTSHOQALDLFQAALPGEPIIRAYADAIYNLTKDPEKK 480  
QY 481 RSLHDYAKKLIQETLLFVDTEHQRPHPMPYLRQVTPESRTKMLDILETLATSKSSD 540  
|  
Db 481 RSLHDYAKKLIQETLLFVDTEHQRPHPMPYLRQVTPESRTKMLDILETLATSKSSD 540  
QY 541 IRLIQLMTEGDANKFPVLAGLLIKIVE 568  
|  
Db 541 IRLIQLMTEGDANKFPVLAGLLIKIVE 568

## RESULT 2

g86493  
probable leader peptide Omp [imported] - Chlamydia pneumoniae (strain J138)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: G86493  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: G86493  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-568 <STO>  
A:Cross-references: GB:BA000008; NID:98978396; PIDN:BA98233.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: Cpj0021

Query Match 100.0%; Score 568; DB 2; Length 568;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGLPFLTFLGGLCSLPISLVAKPESVGHKILYISTOSTQOALATYALDAGDHPF 60  
|  
Db 1 MGLPFLTFLGGLCSLPISLVAKPESVGHKILYISTOSTQOALATYALDAGDHPF 60  
QY 61 VLKRIGEDYLKQSIHSDPQTRKSTIIAGLAGSSEALDVLQAMETADPIQQLVLSAV 120  
|  
b 61 VLKRIGEDYLKQSIHSDPQTRKSTIIAGLAGSSEALDVLQAMETADPIQQLVLSAV 120  
QY 121 SGHIGKTSDDLFRALASPIYIRLEAAYRLANKNTVINDHLSFIKHPLEIOCLSA 180  
|  
Db 121 SGHIGKTSDDLFRALASPIYIRLEAAYRLANKNTVINDHLSFIKHPLEIOCLSA 180  
QY 181 IFLRLTEESDAYIRDLAAKKSARSATLQIGYQOKRFLPTLRNLTSSPDOQAI 240  
|  
Db 181 IFLRLTEESDAYIRDLAAKKSARSATLQIGYQOKRFLPTLRNLTSSPDOQAI 240  
QY 241 LVALGKLKDGOSYNNIKKQKPDVDTLAAQALIALGKEDALPVIKQALERRPAL 300  
|  
Db 241 LVALGKLKDGOSYNNIKKQKPDVDTLAAQALIALGKEDALPVIKQALERRPAL 300  
QY 301 YALRLPEIGIPIALPIFLTKNSEAKLNVALLEIGCDTPKLEXYTERLYOPHYNE 360  
|  
Db 301 YALRLPEIGIPIALPIFLTKNSEAKLNVALLEIGCDTPKLEXYTERLYOPHYNE 360  
QY 361 TLASFSGRGLQNMKRVNIIVPODERERLSTTRGLEQILFELRLPREALPCIT 420  
|  
Db 361 TLASFSGRGLQNMKRVNIIVPODERERLSTTRGLEQILFELRLPREALPCIT 420

QY 421 KLASQKQTALTTAITSFSLTSHOQALDLFQAALPGEPIIRAYADAIYNLTKDPEKK 480  
|  
Db 421 KLASQKQTALTTAITSFSLTSHOQALDLFQAALPGEPIIRAYADAIYNLTKDPEKK 480  
QY 481 RSLHDYAKKLIQETLLFVDTEHQRPHPMPYLRQVTPESRTKMLDILETLATSKSSD 540  
|  
Db 481 RSLHDYAKKLIQETLLFVDTEHQRPHPMPYLRQVTPESRTKMLDILETLATSKSSD 540  
QY 541 IRLIQLMTEGDANKFPVLAGLLIKIVE 568  
|  
Db 541 IRLIQLMTEGDANKFPVLAGLLIKIVE 568

## RESULT 3

g71525  
probable outer membrane leader peptide (omp) CT350 - Chlamydia trachomatis (serotype  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 04-Feb-2000  
C:Accession: G71525  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:9900809  
A:Accession: G71525  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-566 <ARN>  
A:Cross-references: GB:AE001308; GB:AE001273; NID:g3328766; PIDN:AC67945.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: CT350

Query Match 2.1%; Score 12; DB 2; Length 566;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 141 PVIRLEAAYRLA 152  
|  
Db 139 PVIRLEAAYRLA 150

## RESULT 4

conserved hypothetical protein TC0629 [imported] - Chlamydia muridarum (strain N1g9)  
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
C:Accession: A81682  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: A81682  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-566 <TEU>  
A:Cross-references: GB:AE002331; GB:AE002160; NID:g7190662; PIDN:AAF39458.1; PID:g719  
A:Experimental source: strain N1g9 (Mopn)  
C:Genetics:  
A:Gene: TC0629

Query Match 1.8%; Score 10; DB 2; Length 566;  
Best Local Similarity 100.0%; Pred. No. 0.32;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 143 IRLAAYRLA 152  
|  
Db 141 IRLAAYRLA 150

```

RESULT 5
E42409
biphenyl dioxygenase (EC 1.14.-.-) ferredoxin reductase component - Pseudomonas pseudoal
C:Species: Pseudomonas pseudocataliensis
C>Date: 04-Mar-1993 #sequence_revision 06-Jan-1995 #text_change 11-Jun-1999
C:Accession: E42409
R:Taiba, K.; Hirose, J.; Hayashida, S.; Furukawa, K.
J. Biol. Chem. 267, 4844-4853, 1992
A:Title: Analysis of bph operon from the polychlorinated biphenyl-degrading strain of P
A:Reference number: A42409; MUID:92165849
A:Accession: E42409
A:Molecule type: DNA
A:Residues: 1-408 <TAI>
A:Cross-references: GB:M83673; NID:9151090; PIDN:AA25747.1; PID:9151095
A:Experimental source: strain KF707
A>Note: sequence extracted from NCBI backbone (NCBIN:84014, NCBIP:84021)
C:Genetics:
A:Gene: bphA4; bphG
C:Superfamily: toluene dioxygenase ferredoxin reductase component
C:Keywords: aromatic hydrocarbon catabolism; FAD; NAD; oxidoreductase; PCB biodegradatio
F:2-408/Product: biphenyl dioxygenase ferredoxin reductase component #status predicted <
F:9-25/Region: FAD/NAD-binding motif
F:150-166/Region: FAD/NAD-binding motif
F:265-278/Region: FAD/NAD-binding motif

```

```

Query Match 1.6%; Score 9; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 86 ITAGLAGS 94
Db 7 ITAGLAGS 15

```

```

RESULT 6
F41858
biphenyl dioxygenase (EC 1.14.-.-) ferredoxin reductase component - Pseudomonas sp.
C:Species: Pseudomonas sp.
C>Date: 04-Mar-1993 #sequence_revision 06-Jan-1995 #text_change 10-Feb-1995
C:Accession: F41858
R:Ericksen, B.D.; Mondello, F.J.
J. Bacteriol. 174, 2903-2912, 1992
A:Title: Nucleotide sequencing and transcriptional mapping of the genes encoding bipheny
A:Reference number: A41858; MUID:92234948
A:Accession: F41858
A:Molecule type: DNA
A:Residues: 1-408 <ERI>
A:Experimental source: strain LB400
A>Note: sequence extracted from NCBI backbone (NCBIN:97256, NCBIP:97265)
C:Genetics:
A:Gene: bphA4; bphG
C:Superfamily: toluene dioxygenase ferredoxin reductase component
C:Keywords: aromatic hydrocarbon catabolism; FAD; NAD; oxidoreductase; PCB biodegradatio
F:2-408/Product: biphenyl dioxygenase ferredoxin reductase component #status predicted <
F:9-25/Region: FAD/NAD-binding motif
F:150-166/Region: FAD/NAD-binding motif
F:265-278/Region: FAD/NAD-binding motif

```

```

Query Match 1.6%; Score 9; DB 1; Length 408;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 86 ITAGLAGS 94
Db 7 ITAGLAGS 15

```

```

RESULT 7
T03578
probable glucose-inhibited division protein gid - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus

```

```

C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03578
R:Vlock, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SBI
A:Reference number: Z14955; MUID:97404404
A:Accession: T03578
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1445 <VLC>
A:Cross-references: EMBL:AF010496; NID:93128256; PIDN:AC16231.1; PID:93128379
C:Genetics:
A:Map position: 1

```

```

Query Match 1.6%; Score 9; DB 2; Length 445;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 86 ITAGLAGS 94
Db 6 ITAGLAGS 14

```

```

RESULT 8
T05135
hypothetical protein F7H19.220 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05135
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.;
submitted to the Protein Sequence Database, July 1998
A:Reference number: Z15399
A:Accession: T05135
A:Molecule type: DNA
A:Residues: 1-502 <BEV>
A:Cross-references: EMBL:AL031018
A:Experimental source: cultivar Columbia; BAC clone F7H19
C:Genetics:
A:Map position: 4
A>Note: F7H19.220

```

```

Query Match 1.6%; Score 9; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 LLCSLPTS 19
Db 122 LLCSLPTS 130

```

```

RESULT 9
S77785
hypothetical protein MC127 - Mycoplasma capricolum
C:Species: Mycoplasma capricolum
C>Date: 09-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 07-Dec-1999
C:Accession: S77785
R:Bork, P.; Ouzounis, C.; Casari, G.; Schneider, R.; Sander, C.; Dolan, M.; Gilbert,
Mol. Microbiol. 16, 955-967, 1995
A:Title: Exploring the Mycoplasma capricolum genome: a minimal cell reveals its phys
A:Reference number: S77739; MUID:96059641
A:Accession: S77785
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-134 <BOR>
A:Cross-references: EMBL:Z33089; NID:9514453; PIDN:CA83753.1; PID:9530429
A:Experimental source: ATCC 27343
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C:Genetics:
A:Genetic code: SGC3

```

Query Match 1.4%: Score 8; DB 2; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 ITGAGLAG 93  
 |||||  
 Db 7 ITGAGLAG 14

RESULT 10  
 C81695  
 cyclic nucleotide-binding protein, probable TC0506 [imported] - Chlamydia muridarum (str  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000  
 C:Accession: C81695  
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.  
 A:Reference number: A81500; MUID:20150255  
 A:Accession: C81695  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-137 <TEU>  
 A:Cross-references: GB:AE002319; GB:AE002160; NID:g7190541; PIDN:AAE39348.1; PID:g719054  
 A:Experimental source: strain N1gg (Mopn)  
 C:Genetics:  
 A:Gene: TC0506

Query Match 1.4%: Score 8; DB 2; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 331 VALALLEL 338  
 |||||  
 Db 120 VALALLEL 127

RESULT 11  
 E95203  
 conserved hypothetical protein SPI746 [imported] - Streptococcus pneumoniae (strain TIG  
 C:Species: Streptococcus pneumoniae  
 C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
 C:Accession: E95203  
 R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei  
 nson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A:Reference number: A95000; MUID:21357209; PMID:11463916  
 A:Accession: E95203  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-197 <KUR>  
 A:Cross-references: GB:AE005672; PIDN:AAK75822.1; PID:g14973242; GSPDB:GN00164; TIGR:SP4  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SPI746  
 C:Superfamily: hypothetical protein ygek

Query Match 1.4%: Score 8; DB 2; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 LHDYAKKL 490  
 |||||  
 Db 57 LHDYAKKL 64

RESULT 12

E98070  
 conserved hypothetical protein spi1591 [imported] - Streptococcus pneumoniae (strain  
 C:Species: Streptococcus pneumoniae  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C:Accession: E98070  
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: E98070  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-197 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAU00394.1; PID:g15459258; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: spi1591  
 C:Superfamily: hypothetical protein ygek

Query Match 1.4%: Score 8; DB 2; Length 197;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 483 LHDYAKKL 490  
 |||||  
 Db 57 LHDYAKKL 64

RESULT 13  
 A12601  
 conserved hypothetical protein Atu0208 [imported] - Agrobacterium tumefaciens (strain  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: A12601  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: A12601  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-308 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AAU41231.1; PID:g17738535; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu0208  
 A:Map position: circular chromosome

Query Match 1.4%: Score 8; DB 2; Length 308;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 ITGAGLAG 93  
 |||||  
 Db 189 ITGAGLAG 196

RESULT 14  
 A97384  
 hypothetical protein AGR\_C357 [imported] - Agrobacterium tumefaciens (strain C58, Ce  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
 C:Accession: A97384  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001  
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
 A:Reference number: A97359; PMID:11743194  
 A:Accession: A97384  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-328 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK86026.1; PID:g15155095; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR.C.357  
 A:Map position: circular chromosome

Query Match 1.4%; Score 8; DB 2; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 IIGAGIAG 93  
 |||||  
 Db 209 IIGAGIAG 216

RESULT 15  
 T03557  
 ribose transport system permease protein RbSC - *Rhodobacter capsulatus*  
 C:Species: *Rhodobacter capsulatus*  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 24-Nov-1999  
 C:Accession: T03557  
 R:Vleck, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997  
 A>Title: Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1003  
 A:Reference number: Z14955; MUID:97404404  
 A:Accession: T03557  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-357 <VLC>  
 A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAQ16210.1; PID:g3128358  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: 1-arabinose transport system permease arah  
 C:Keywords: ribose transport

Query Match 1.4%; Score 8; DB 2; Length 357;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 LAAAGALI 276  
 |||||  
 Db 86 LAAAGALI 93

Search completed: August 20, 2002, 04:26:37  
 Job time: 256 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:26:01 ; Search time 17.08 Seconds

(Without alignments)  
1287.629 Million cell updates/sec

Title: US-09-662-812-2

Sequence: 1 MCLFHLTFGLLCSLPISL.....TEGDAKNPVLAGLLIKIVE 568

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.6	408	1	BPHG_BURCE
2	8	1.4	342	1	SNXF_HUMAN
3	8	1.4	435	1	GID_BACSU
4	8	1.4	447	1	ASTB_ECOLI
5	8	1.4	558	1	ZDS_ARATH
6	8	1.4	570	1	ZDS_MATZE
7	8	1.4	574	1	ZDS_MARPS
8	8	1.4	587	1	ZDS_TARER
9	8	1.4	588	1	ZDS_CAPAN
10	8	1.4	588	1	ZDS_LYCES
11	8	1.4	666	1	FLDP_VIBPA
12	8	1.4	761	1	RIRI_BUCAI
13	8	1.2	85	1	GONZ_DICLA
14	7	1.2	85	1	GONZ_MORSA
15	7	1.2	106	1	VMEK_CVB
16	7	1.2	122	1	SAAL_RABIT
17	7	1.2	122	1	SSAM_SALTY
18	7	1.2	137	1	RT16_HUMAN
19	7	1.2	147	1	PA24_BUNMU
20	7	1.2	148	1	YEES_ECOLI
21	7	1.2	163	1	LEUD_PYRHO
22	7	1.2	164	1	RECK_TREPA
23	7	1.2	186	1	YJ89_MYCTU
24	7	1.2	187	1	NFX8_PSEAE
25	7	1.2	207	1	GL21_ARATH
26	7	1.2	213	1	KGUA_CAUCR
27	7	1.2	230	1	FLAH_METVO
28	7	1.2	259	1	YCBC_ECOLI
29	7	1.2	260	1	ATP6_BRANA
30	7	1.2	286	1	PUR7_PASMU
31	7	1.2	291	1	ATP6_MAIZE
32	7	1.2	306	1	PUR7_HAEIN
33	7	1.2	306	1	P43851 haemophilus

34	7	1.2	306	1	YDS4_SCHPO	014180 schizosach
35	7	1.2	309	1	HCD2_CAEEL	P41938 caenorhabd1
36	7	1.2	321	1	EUM1_EUMMA	P25780 euroglyphus
37	7	1.2	324	1	PHT2_PSEPU	005182 pseudomonas
38	7	1.2	329	1	DONS_MOUSE	09qxp4 mus musculu
39	7	1.2	331	1	IPNS_PENCH	P08703 penicillium
40	7	1.2	332	1	Y675_TREPA	083681 treponema p
41	7	1.2	371	1	CYB_ASPEE	048017 aspidites m
42	7	1.2	371	1	CYB_LIALI	048090 liasis albe
43	7	1.2	371	1	CYB_LIACH	048092 liasis albe
44	7	1.2	371	1	CYB_LIAMA	048093 liasis macu
45	7	1.2	371	1	CYB_LIAMS	048094 liasis mack

## ALIGNMENTS

RESULT 1  
BPHG\_BURCE STANDARD; PRT; 408 AA.  
ID BPHG\_BURCE  
AC P37337;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE Biphenyl dioxygenase system ferredoxin--NAD(+) reductase component  
(EC 1.18.1.3).  
GN BPHG.  
OS Burkholderia cepacia (Pseudomonas cepacia).  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Burkholderia.  
OX NCBI\_Taxid:292;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LB400;  
RX MEDLINE=92234948; Pubmed=1569021;  
RA Erickson B.D., Mondello F.J.;  
RT "Nucleotide sequencing and transcriptional mapping of the genes  
RT encoding biphenyl dioxygenase, a multicomponent  
RT polychlorinated biphenyl-degrading enzyme in Pseudomonas strain  
RT LB400".  
RT J. Bacteriol. 174:2903-2912(1992).  
RN [2]  
RP REVISIONS.  
RA Erickson B.D., Mondello F.J.  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
CC - FUNCTION: PART OF THE ELECTRON TRANSFER COMPONENT OF BIPHENYL  
CC DIOXYGENASE. TRANSFERS ELECTRONS FROM FERREDOXIN (BPHF) TO  
CC NADH.  
CC - CATALYTIC ACTIVITY: Reduced ferredoxin + NAD(+) = oxidized  
CC ferredoxin + NADH.  
CC - COFACTOR: FAD.  
CC - PATHWAY: BIPHENYL-POLYCHLORINATED BIPHENYL DEGRADATION PATHWAY.  
CC - SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE  
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA AND BPHE), A  
CC FERREDOXIN (BPHF) AND A FERREDOXIN REDUCTASE (BPHG).  
CC - SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE  
CC FERREDOXIN REDUCTASE COMPONENTS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M86348; AAB63429.1; -  
CC PIR: F41858.  
CC InterPro: IPR001327; FAD\_Pyr\_redox.  
CC InterPro: IPR002025; NAD\_binding.  
CC Pfam: PF00070; Pyr\_redox; 1.  
CC DR PRINTS: PR00368; FADPNR.  
CC Aromatic hydrocarbons catabolism; Flavoprotein; FAD; NAD;

KW Oxidoreductase.  
 FT NP\_BIND 4 35 FAD (ADP PART) (POTENTIAL).  
 ET NP\_BIND 145 173 NAD (ADP PART) (POTENTIAL).  
 SO SEQUENCE 408 AA; 42953 MW; 8A52BB01688667A9 CRC64;

Query Match 1.6%; Score 9; DB 1; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGIAGS 94  
 DB 7 IIGAGIAGS 15

## RESULT 2

SNXF\_HUMAN SNXF\_HUMAN STANDARD; PRT; 342 AA.

AC Q9NRS6; Q9NRS5;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sorting nexin 15.

GN SNX15.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RA PHILLIPS S.A., Barr V.A., Haft D.H., Taylor S.I., Haft C.R.;

RC MEDLINE=21265032; PubMed=11085978;

RT "Identification and characterization of snx15, a novel sorting nexin

involved in protein trafficking.";

RL J. Biol. Chem. 276:5074-5084(2001).

CC -1- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR

TRAFFICKING. OVEREXPRESSION OF SNX15 DISRUPTS THE NORMAL

TRAFFICKING OF PROTEINS FROM THE PLASMA MEMBRANE TO RECYCLING

ENDOSOMES OR THE TGN.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2/SNX15A;

CC ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.

CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.

CC -1- SIMILARITY: CONTAINS 1 PHOX HOMOLOG (PX) DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: AF175267; AAF89955.1; -

DR EMBL: AF175268; AAF89956.1; -

DR MIM: 605964; -

DR InterPro: IPR001683; PX.

DR Pfam: PF00787; PX; 1.

DR SMART: SM00312; PX; 1.

KW Transport; Protein transport; Alternative splicing.

FT DOMAIN 11 123 PX.

FT VARSPLIT 221 308

FT EEGAAPSPTVAALATMEVSARLDQEPWPGQEEEDGE

FT GGPPAYLSQATELITQALNDEKAGAYAAALOGYRDGVHVL

FT LOGVPS -> G (IN ISOFORM 2).

SEQUENCE 342 AA; 38291 MW; 33F64A79EAF6BDBC CRC64;

Query Match 1.4%; Score 8; DB 1; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 444 QEAIDLLE 451

|||||||  
 143 QEAIDLLE 140

## RESULT 3

GID\_BACSU GID\_BACSU STANDARD; PRT; 435 AA.

AC P39815;

DT 01-FEB-1995 (Rel. 31, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE protein gid.

GN GID.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI\_TaxID=1423;

RA [1]

RC SEQUENCE FROM N.A.

DT STRAIN=168;

DT Foulger D., Errington J.;

RT "Cloning and sequencing 7.5Kbp of DNA from Bacillus subtilis upstream

of the codv gene.";

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

CC [2]

CC SEQUENCE OF 1-363 FROM N.A.

CC STRAIN=168 / 8G5;

CC de Jong S.;

RT "Cloning and sequencing of the TopI gene, the gene encoding B.

subtilis DNA topoisomerase I.";

CC Submitted (XXX-1994) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE GIDA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL: AJ000975; CA044423.1; -

DR EMBL: Z99112; CAB13486.1; -

DR EMBL: L27797; AAA22764.1; -

DR Subtilist; BG11008; gid.

DR InterPro: IPR002218; GIDA.

DR InterPro: IPR00205; NAD binding.

DR PROSITE: PS01280; GIDA\_1; FALSE\_NEG.

DR PROSITE: PS01281; GIDA\_2; 1.

KW Complete proteome.

SEQUENCE 435 AA; 48063 MW; 6C7AB028F484B683 CRC64;

OY 87 IIGAGIAGS 94

DB 9 IIGAGIAGS 16

## RESULT 4

ASTB\_ECOLI ASTB\_ECOLI STANDARD; PRT; 447 AA.

AC P76216;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Succinylarginine dihydrolase (EC 3.-.-.-).

GN ASTB OR B1745.

OS Escherichia coli.



OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426517; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RL "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN (2)  
 RP FUNCTION.  
 RX MEDLINE=98361920; PubMed=9696779;  
 RA Schneider B.L., Knapakis A.K., Reitzer L.J.;  
 RA "Arginine catabolism and the arginine succinyltransferase pathway in  
 RA Escherichia coli.";  
 RT RT  
 CC -1- FUNCTION: CATALYZES THE HYDROLYSIS OF N(2)-SUCCINYLRARGININE INFO  
 CC N(2)-SUCCINYLRORITHINE, AMMONIA AND CO(2).  
 CC -1- PATHWAY: SECOND STEP IN ARGININE CATABOLISM BY THE ARGININE  
 CC SUCCINYLRNTRANSFERASE PATHWAY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF000269; AAC74815.1; -  
 DR ECocore; EC13396; asfB.  
 KW Arginine metabolism; Hydrolase; Complete proteome.  
 SQ SEQUENCE 447 AA; 49298 MW; 710AE691E413AF64 CRC64;  
 Query Match 1.4%; Score 8; DB 1; Length 447;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 96 EALDVLSQ 103  
 Db 422 EALDVLSQ 429  
 RESULT 5  
 ZDS\_ARATH STANDARD; PRT; 558 AA.  
 AC 038693: Q9LLYL; Q9CAV3;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)  
 DE (Carotene 7,8-desaturase).  
 GN ZDS1 OR ZDS OR AT3G04870 OR T9J14.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Scolnik P.A., Bartley G.E.;  
 RT "Nucleotide sequence of zeta-carotene desaturase from Arabidopsis.";  
 RL (in) Plant Gene Register PGR95-111.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RA Giuliano G., Rosati C., Santangelo G.;  
 RT "Gene structure and regulation of the carotenoid biosynthetic pathway

RT in Arabidopsis.";  
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016720; PubMed=11130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerg W., Unsel M.,  
 RA Fartman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Deleney M., Boutry M., Grivell L.A., Maché R., Puigdomenech P.,  
 RA de Simone V., Choise N., Attienave F., Robert C., Brotier P.,  
 RA Wincker P., Cattolico L., Weissenbach J., Saurin W., Queller F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,  
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Yakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,  
 RA Reichelt J., Schafte M., Schoen O., Barques M., Terol J., Climent J.,  
 RA Navarro P., Collado C., Perez-Perez A., Ottenwelder B., Duchemin D.,  
 RA Cooke R., Laudie M., Berger-Llauró C., Purnelle B., Masuy D.,  
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 RA Monfort A., Argirou A., Flores M., Iguori R., Vitale D.,  
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,  
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,  
 RA Preuss D., Lin X., Nieman W.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:820-822(2000).  
 CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene  
 CC via the intermediary of neurosporene. It carries out two  
 CC consecutive desaturations (introduction of double bonds) at  
 CC positions C-7 and C-7'.  
 CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +  
 CC A + 2 H(2)O.  
 CC -1- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2  
 CC H(2)O.  
 CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).  
 CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.  
 CC -1- SUBCELLULAR LOCATION: Chloroplast; Chromoplast.  
 CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U38550; AAA91161.1; -  
 DR EMBL: AF121947; AAF85796.1; -  
 DR EMBL: AC009465; AAG51402.1; -  
 DR Mendel: 581; Arath; Zds1; 581.  
 DR InterPro: IPR000759; Adirndx\_reductase.  
 DR InterPro: IPR000205; NAD\_binding.  
 DR PRIN1: PR00419; ADXRPTASE.  
 KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;  
 KM Chloroplast; Transit peptide.  
 FT TRANSIT 1 27  
 FT CHAIN 28 558  
 FT CONFLICT 9 9 A -> P (IN REF. 1).  
 FT CONFLICT 93 93 G -> A (IN REF. 1).  
 FT CONFLICT 172 172 N -> S (IN REF. 1).  
 FT CONFLICT 370 370 G -> V (IN REF. 3).  
 FT CONFLICT 463 463 S -> P (IN REF. 1).  
 FT CONFLICT 470 470 C -> W (IN REF. 3).

SQ SEQUENCE 558 AA; 61508 MW; FDE5AE5DFB4A531D CRC64;

Query Match 1.4%; Score 8; DB 1; Length 558;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93  
| | | | | | | |  
DB 61 IIGAGLAG 68

## RESULT 6

ZDS\_MAIZE  
ID ZDS\_MAIZE STANDARD; PRT; 570 AA.  
AC Q9TPT4;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)  
DE (Carotene 7,8-desaturase).  
GN ZDS1 OR ZDS.  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Rosaceae; PACC clade;  
OC Panicoidae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Luo R., Wurtzel E.T.;  
RT "A maize cDNA encoding zeta carotene desaturase."  
RL (In) Plant Gene Register PGR99-118.  
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene  
via the intermediary of neosporene. It carries out two  
consecutive desaturations (introduction of double bonds) at  
positions C-7 and C-7'.  
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neosporene +  
A + 2 H(2)O.  
CC -1- CATALYTIC ACTIVITY: Neosporene + AH(2) + O(2) = lycopene + A + 2  
H(2)O.  
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).  
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.  
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF047490; AAD02462.1; -;  
DR Mendel; 35948; Zeama; zds1.35948.  
DR InterPro: IPR000759; Adnrx\_reductase.  
DR InterPro: IPR003042; Rng\_moxxygenase.  
DR PRINTS; PR00419; ADXRDTASE.  
DR PRINTS; PR00420; RNCMNOMXGNASE.  
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;  
KW Chloroplast; Transl. peptide.  
FT TRANSIT 1 570 ZETA-CAROTENE DESATURASE.  
FT CHAIN ? 570 ZETA-CAROTENE DESATURASE.  
SQ SEQUENCE 570 AA; 63127 MW; C7ADB2FD62531E8 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 570;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93  
| | | | | | | |  
DB 69 IIGAGLAG 76

RESULT 7  
ZDS\_NARPS  
ID ZDS\_NARPS STANDARD; PRT; 574 AA.  
AC O49901;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)  
DE (Carotene 7,8-desaturase).  
GN ZDS1.  
OS Narcissus pseudonarcissus (Daffodil).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;  
OC Narcissus.  
OX NCBI\_TaxID=39639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Al-Babili S., Oelschlegel J., Beyer P.;  
RT "A cDNA encoding for beta carotene desaturase from Narcissus  
pseudonarcissus L.";  
RL (In) Plant Gene Register PGR98-103.  
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene  
via the intermediary of neosporene. It carries out two  
consecutive desaturations (introduction of double bonds) at  
positions C-7 and C-7'.  
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neosporene +  
A + 2 H(2)O.  
CC -1- CATALYTIC ACTIVITY: Neosporene + AH(2) + O(2) = lycopene + A + 2  
H(2)O.  
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).  
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.  
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AJ224683; CAA12062.1; -;  
DR Mendel; 28308; Narps; zds1.28308.  
DR InterPro: IPR002937; Amino-oxidase.  
DR Pfam: PF01593; Amino-oxidase; 1.  
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;  
KW Chloroplast; Transl. peptide.  
FT TRANSIT 1 574 CHLOROPLAST (POTENTIAL).  
FT CHAIN ? 574 ZETA-CAROTENE DESATURASE.  
SQ SEQUENCE 574 AA; 63581 MW; 67D2C8DAFF942F7 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 574;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93  
| | | | | | | |  
DB 69 IIGAGLAG 76

## RESULT 8

ZDS\_TARER  
ID ZDS\_TARER STANDARD; PRT; 587 AA.  
AC Q9PTV46;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)  
DE (Carotene 7,8-desaturase).  
GN ZDS1.  
OS Narcissus pseudonarcissus (Daffodil).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;  
OC Narcissus.  
OX NCBI\_TaxID=39639;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Al-Babili S., Oelschlegel J., Beyer P.;  
RT "A cDNA encoding for beta carotene desaturase from Narcissus  
pseudonarcissus L.";  
RL (In) Plant Gene Register PGR98-103.  
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene  
via the intermediary of neosporene. It carries out two  
consecutive desaturations (introduction of double bonds) at  
positions C-7 and C-7'.  
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neosporene +  
A + 2 H(2)O.  
CC -1- CATALYTIC ACTIVITY: Neosporene + AH(2) + O(2) = lycopene + A + 2  
H(2)O.  
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).  
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.  
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.  
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AJ224683; CAA12062.1; -;  
DR Mendel; 28308; Narps; zds1.28308.  
DR InterPro: IPR002937; Amino-oxidase.  
DR Pfam: PF01593; Amino-oxidase; 1.  
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;  
KW Chloroplast; Transl. peptide.  
FT TRANSIT 1 574 CHLOROPLAST (POTENTIAL).  
FT CHAIN ? 574 ZETA-CAROTENE DESATURASE.  
SQ SEQUENCE 574 AA; 63581 MW; 67D2C8DAFF942F7 CRC64;

```

OS Tagetes erecta (African marigold).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
OC Asteridae: euasterids II: Asterales: Asteraceae: Asteroideae:
OC Helianthus: Tagetes.
OX NCBI_TaxID=13708;
RN [1]
RP SEQUENCE FROM N.A.
RA Moehs C.P., Tian L., Dellapenna D.;
RT "Analysis of carotenoid biosynthetic gene expression during marigold
RT petal development.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
CC via the intermediary of neurosporene. It carries out two
CC consecutive desaturations (introduction of double bonds) at
CC positions C-7 and C-7' (By similarity).
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
CC A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
CC H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF251013; AAG10425.1; -.
DR InterPro: IPR000759; Adnrxn_reductase.
DR InterPro: IPR003042; Rng_omoxigenase.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00420; RNMNMOXGNASE.
KM Carotenoid biosynthesis: Oxidoreductase; NAD; Flavoprotein; FAD;
KM Chloroplast; Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 587 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 587 AA; 65085 MW; 34D2DF150729B99C CRC64;

Query Match 1.4%; Score 8; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITGAGLAG 93
Db 87 ITGAGLAG 94

```

```

RA Albrecht M., Klein A., Hugueney P., Sandmann G., Kuntz M.;
RT "Molecular cloning and functional expression in E. coli of a novel
RT plant enzyme mediating zeta-carotene desaturation.";
RL FEBS Lett. 372:199-202(1995).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99421754; PubMed=10491195;
RA Albrecht M., Klein A., Hugueney P., Sandmann G., Kuntz M.;
RT "Catalytic properties of an expressed and purified higher plant type
RT zeta-carotene desaturase from Capsicum annuum.";
RL Eur. J. Biochem. 265:376-383(1999).
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
CC via the intermediary of neurosporene. It carries out two
CC consecutive desaturations (introduction of double bonds) at
CC positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
CC A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
CC H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X89897; CA61985.1; -.
DR InterPro: IPR000759; Adnrxn_reductase.
DR InterPro: IPR003042; Rng_omoxigenase.
DR PRINTS: PR00419; ADXRDTASE.
DR PRINTS: PR00420; RNMNMOXGNASE.
KM Carotenoid biosynthesis: Oxidoreductase; NAD; Flavoprotein; FAD;
KM Chloroplast; Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 588 ZETA-CAROTENE DESATURASE.
SQ SEQUENCE 588 AA; 64684 MW; 55F568FAEE7EA91 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 588;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 ITGAGLAG 93
Db 84 ITGAGLAG 91

```

```

RESULT 9
ZDS_CAPAN
ID ZDS_CAPAN STANDARD; PRT; 588 AA.
AC Q9SMJ3;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, last sequence update)
DT 01-MAR-2002 (Rel. 41, last annotation update)
DE zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
DE (Carotene 7,8-desaturase).
GN ZDS.
OS Capsicum annuum (Bell pepper).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
OC Asteridae: euasterids I; Solanales: Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV, LAMOYO; TISSUE=fruit;
RX MEDLINE=96000207; PubMed=7556669;

```

```

RESULT 10
ZDS_LYCES
ID ZDS_LYCES STANDARD; PRT; 588 AA.
AC Q9SE20;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, last sequence update)
DT 01-MAR-2002 (Rel. 41, last annotation update)
DE zeta-carotene desaturase, chloroplast precursor (EC 1.14.99.30)
DE (Carotene 7,8-desaturase).
GN ZDS.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
OC Asteridae: euasterids I; Solanales: Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA Bartley G.E., Ishida B.K.;
RT "Zeta-carotene desaturase from tomato.";

```

```

RL (In) Plant Gene Register PCR99-181.
CC -1- FUNCTION: Catalyzes the conversion of zeta-carotene to lycopene
CC via the intermediary of neurosporene. It carries out two
CC consecutive desaturations (introduction of double bonds) at
CC positions C-7 and C-7'.
CC -1- CATALYTIC ACTIVITY: zeta-carotene + AH(2) + O(2) = neurosporene +
CC A + 2 H(2)O.
CC -1- CATALYTIC ACTIVITY: Neurosporene + AH(2) + O(2) = lycopene + A + 2
CC H(2)O.
CC -1- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -1- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -1- SIMILARITY: BELONGS TO THE ZETA CAROTENE DESATURASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF195507; AAF13698.1; -.
CC InterPro: IPR000759; Adrxn_reductase.
CC InterPro: IPR003042; Rng_ammoxigenase.
CC PRINTS: PR00419; AOXRDASE.
CC PRINTS: PR00420; RINGMONOXINASE.
CC Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
CC Chloroplast; Transit peptide.
CC TRANSLIT ? 588 ZETA-CAROTENE DESATURASE.
CC CHAIN 1 ? 588 CHLOROPLAST (POTENTIAL).
CC FT SEQUENCE 588 AA; 64733 MW; 786675800C1B846 CRC64;
CC -----
Query Match 1.4%; Score 8; DB 1; Length 588;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 IIGAGLAG 93
DB 84 IIGAGLAG 91

```

```

CC A SINGLE POLAR FLAGELLUM PROPELS THE BACTERIUM IN LIQUID
CC (SWIMMING), WHILE MULTIPLE LATERAL (PERTURCHOUS) FLAGELLA MOVE
CC THE BACTERIUM OVER SURFACES (SWIMMING). THE POLAR FLAGELLUM IS
CC SYNTHESIZED CONSTITUTIVELY BUT LATERAL FLAGELLA ARE PRODUCED ONLY
CC UNDER CONDITIONS IN WHICH THE POLAR FLAGELLUM IS NOT FUNCTIONAL.
CC -1- SIMILARITY: BELONGS TO THE FLID FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF069392; AAC27803.1; -.
CC InterPro: IPR003481; FLID.
CC Pfam: PF02465; FLID. 1.
CC K1 FLAGELLA; Coiled coil.
CC FT DOMAIN 338 424 COILED COIL (POTENTIAL).
CC SEQUENCE 666 AA; 72761 MW; EA2CA623ADA69F8A CRC64;
CC -----
Query Match 1.4%; Score 8; DB 1; Length 666;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 385 DPOERERL 392
DB 395 DPOERERL 402

```

```

FT PEPTIDE 24 33 GONADOLIBERIN II.
FT PEPTIDE 37 85 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID
FT MOD_RES 33 33 (BY SIMILARITY).
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
FT SIMILARITY).
SQ SEQUENCE 85 AA; 9646 MW; F832C0698C842C64 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LEGLLC 14
|||||||
Db 9 LFGLLLC 15

RESULT 14
GON2_MORSA STANDARD; PRT; 85 AA.
AC 073811.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)
DE (GNRH-II) (LH-RH II) (Luliberin II).
GN GNRH2.
OS Morone saxatilis (Striped bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Moronidae; Morone.
ON NCBI_TaxID=34816;
[1]
RX SEQUENCE FROM N.A.
RA Chow M.M., Kight K.E., Gotthilf Y., Alok D., Zohar Y.;
RT "Multiple GnRHs present in a teleost species are encoded by separate
RT genes: analysis of the sbGNRH and cGNRH-II genes from the striped
RT bass, Morone saxatilis. ";
RL submitted (MAR-1998) to the EMBL/GenBank/DDNJ databases.
CC -!- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF056313; AAD03816.1; -
DR InterPro; IPR002012; GNRH.
DR Pfam; PF00446; GNRH.1.
DR PROSITE; PS00473; GNRH.1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 85 PROCONADOLIBERIN II.
FT PEPTIDE 24 33 GONADOLIBERIN II.
FT PEPTIDE 37 85 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID
FT MOD_RES 33 33 (BY SIMILARITY).
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
FT SIMILARITY).
SQ SEQUENCE 85 AA; 9673 MW; F832D5B3BC842C64 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 1.2%; Score 7; DB 1; Length 85;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LFGLLC 14  
 |||||  
 Db 9 LFGLLC 15

## RESULT 15

VMEM\_CVB STANDARD; PRT; 106 AA.  
 ID VMEM\_CVB  
 AC P37989;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 11.4 kDa membrane protein (ORF 3).  
 OS Chrysanthemum virus B (CVB).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.  
 OX NCBI\_TaxID=12165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92013948; PubMed=1919520;  
 RA LeVay K., Zavrlev S.;  
 RT "Nucleotide sequence and gene organization of the 3'-terminal region  
 of chrysanthemum virus B genomic RNA."  
 RL J. Gen. Virol. 72:2333-2337(1991).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: TO ORF3 PROTEIN FROM POTEXVIRUSES AND TO THE 14 kDa  
 CC PROTEIN FROM BSMV RNA 2BETA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; S60150; AAB20078.2; -.  
 DR PIR; J01248; J01248.  
 DR InterPro; IPR001896; Plant\_vir\_prot.  
 DR Pfam; PF01307; Plant\_vir\_prot; 1.  
 DR ProDom; PD001561; Plant\_vir\_prot; 1.  
 KW Transmembrane.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 72 92 POTENTIAL.  
 SO SEQUENCE 106 AA; 11435 MW; B3366CFBB40C92D6 CRC64;

Query Match 1.2%; Score 7; DB 1; Length 106;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 177 LSAIFL 183  
 |||||  
 Db 85 LSAIFL 91

Search completed: August 20, 2002, 04:30:39  
 Job time: 278 sec

Run on: August 20, 2002, 01:59:15 ; Search time 220.03 seconds  
(without alignments)

14445.867 Million cell updates/sec

US-09-662-812-1  
1907

Perfect score: 1507  
Sequence: 1 gtggttgatttgaagaag.....ctcatcaagaggctgtgaa 1507

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters:

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

```
Database : N_Geneseq_032802:*
1: /SIPI1/qcdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
```

1: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1907	100.0	1907	22	AA003024	Chlamydia pneumoniae
2	24	1.3	36	22	AA003026	Chlamydia pneumoniae
3	24	1.3	44	22	AA003025	Chlamydia pneumoniae
4	21	1.1	433	21	AA075007	Nucleotide sequence
5	20	1.0	20	20	AA094540	PCR primer used to
6	20	1.0	20	20	AA094538	PCR primer used to
7	20	1.0	20	20	AA092063	PCR primer used to
8	20	1.0	20	20	AA092065	PCR primer used to
9	20	1.0	20	20	AA092068	PCR primer used to

C	10	20	1.0	20	20	AA929070
C	11	19	1.0	800	20	AA116609
C	12	19	1.0	1210	22	AA070852
C	13	19	1.0	1665	23	AA564487
C	14	19	1.0	1958	22	ABA09201
C	15	19	1.0	1958	22	AA988573
C	16	19	1.0	2239	23	AAAD3450
C	17	19	1.0	5981	23	AAAD3461
C	18	19	1.0	6062	13	AA020765
C	19	19	1.0	6062	13	AA020782
C	20	19	1.0	6414	11	AA006283
C	21	19	1.0	6414	13	AAQ02766
C	22	19	1.0	6414	13	AAQ02733
C	23	19	1.0	10719	23	ABL2106
C	24	19	1.0	37808	20	AAK02780
C	25	19	1.0	55583	21	AAFE2125
C	26	19	1.0	55583	21	AAA35003
C	27	18	0.9	138	21	AAA68058
C	28	18	0.9	207	21	AAA65034
C	29	18	0.9	401	21	AAA77766
C	30	18	0.9	401	22	AA128504
C	31	18	0.9	512	22	AAK76246
C	32	18	0.9	512	22	AAK76247
C	33	18	0.9	512	22	AAK76248
C	34	18	0.9	755	22	AA194747
C	35	18	0.9	755	22	AA194748
C	36	18	0.9	759	22	AAK91824
C	37	18	0.9	780	22	AA195498
C	38	18	0.9	794	22	AAK92007
C	39	18	0.9	794	22	AAK93667
C	40	18	0.9	880	19	AA596234
C	41	18	0.9	904	22	AA193979
C	42	18	0.9	1020	18	AAK30764
C	43	18	0.9	1030	19	AA798656
C	44	18	0.9	1464	21	AAAB69017
C	45	18	0.9	1723	5	AAAD0246

## ALIGNMENTS

RESULT 1

AAD03024 standard; DNA; 1907 BP.

AA  
AC AAD03024;

AA DT 13-JUN-2001 (first entry)

XX chlamydia pneumoniae outer membrane protein (OMP) gene

XX outer membrane protein; OMP; antibacterial; vaccine; gene therapy;

antibody; medicament; Chlamydia infection; ds.

Chlamydia pneumoniae

AA FH	Key	Location/Qualifiers
AA		1007
FH		1007

CDS  
FTT  
FTT

/\*tag= a  
tag= a

outer membrane protein (OMP)"

Et

PN WO200121804-A1

AA PD 29-MAR-2001

AA  
PF 15-SEP-2000; 2000WO-CA01088.

XX 20-SEP-1999; 99US-0154652.  
PR

XX  
PA (AVET ) AVENTIS PASTEUR LTD.

XX Oomen RP, Wang J, Dunn P:  
PI Mordin AD,

XX

PCR primer used to  
Human gene express  
Human secreted pro  
DNA encoding novel  
Human cysteine/Glu  
Human EST-derived  
Human lung tumour-  
Human lung tumour-  
pAD-CMV19, AAQ20  
pAD-CMV19, express1  
pAD-CMV1 confg, a  
pAD-CMV1, AAQ207  
pAD-CMV1, express2  
Drosophila melanog  
vector, pMYC-BG DNA  
Human low adenosis re  
Human adenosis re  
Bactetiophage 44A8  
Bactetiophage 44A8  
cDNA encoding huma  
Colon tumour relat  
Human immune/haema  
Human immune/haema  
Human immune/haema  
Human neuroblastom  
Human cDNA 5'-end  
Human cDNA clone 1  
Human neuroblastom  
Human cDNA 5'-end  
Human cDNA clone 1  
S. pneumoniae del $\Delta$   
Human neuroblastom  
Streptococcus pneu  
DNA encoding a S.  
Bacteriophage 44A1  
DNA segment from p

DR WPI: 2001-244939/25.  
 DR P-PSDB: AAT72972.

PT Novel Chlamydia pneumoniae outer membrane protein and polynucleotides  
 PT encoding them, useful as components of vaccines for treating Chlamydia  
 PT infections, and for detecting Chlamydia infections in the body fluids  
 of mammals -  
 XX  
 XX  
 PS  
 PS

Claim 2; Fig 1; 82pp; English.

The present sequence is a DNA encoding Chlamydia pneumoniae OMP (outer membrane protein). OMP is useful in the production of vaccines, antibodies and pharmaceutical compositions which are useful for treating or preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum) infections. They are also useful as diagnostic reagents for detecting Chlamydia infection which involves assaying a body fluid of a mammal to be tested for the components. The OMP vaccine is useful in the preparation of a medicament for preventing and/or treating Chlamydia infection. The primers derived from OMP gene are also useful for detecting and/or identifying Chlamydia in a biological material. OMP antibodies are also useful as reagents for purifying OMP from a biological sample which involves carrying out antibody-based affinity chromatography with the biological sample. OMP gene is also useful in gene therapy.

Sequence 1907 BP; 542 A; 456 C; 380 G; 529 T; 0 other;

Query Match 100.0%; Score 1907; DB 22; Length 1907;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtggtgattcttcaaaaaggtccatgagtggttataatgttcaagcttcctatcc 60  
 Db 1 gtggtgattcttcaaaaaggtccatgagtggttataatgttcaagcttcctatcc 60  
 QY 61 aaacatgaaatctctagagagtgatgaaatcgatcctatgagagctatccatcaac 120  
 Db 61 aaacatgaaatctctagagagtgatgaaatcgatcctatgagagctatccatcaac 120  
 QY 121 tctcttgagcttatttctgttagtcttccattctctgttctgttcaattccctgagtc 180  
 Db 121 tctcttgagcttatttctgttagtcttccattctctgttctgttcaattccctgagtc 180  
 QY 181 tctcttgagcttatttctgttagtcttccattctctgttctgttcaattccctgagtc 180  
 Db 181 tctcttgagcttatttctgttagtcttccattctctgttctgttcaattccctgagtc 180  
 QY 241 tctggaagctctagatgctacggtgacatgacttctctgtttaaagaaatcggaga 300  
 Db 241 tctggaagctctagatgctacggtgacatgacttctctgtttaaagaaatcggaga 300  
 QY 301 agactatctcaagcaaaagcatcactcctcagatccgcaactagaaagacacatcat 360  
 Db 301 agactatctcaagcaaaagcatcactcctcagatccgcaactagaaagacacatcat 360  
 QY 361 tggaaagagctggtgagatctcagaagccttgagatgcttccccaagtctggaagc 420  
 Db 361 tggaaagagctggtgagatctcagaagccttgagatgcttccccaagtctggaagc 420  
 QY 421 tgcagacccctcagagagctggttattatcggaagctcagagacatctgggaaagc 480  
 Db 421 tgcagacccctcagagagctggttattatcggaagctcagagacatctgggaaagc 480  
 QY 481 ttcgagacactactgtttaagctttaagatcctccttctctgtcactcgttagaagc 540  
 Db 481 ttcgagacactactgtttaagctttaagatcctccttctctgtcactcgttagaagc 540  
 QY 541 cgcctatagcttgtaatttgaagaacactaaagtcatttgaatcactacattcttcat 600  
 Db 541 cgcctatagcttgtaatttgaagaacactaaagtcatttgaatcactacattcttcat 600  
 QY 601 tcatagcttcccgagaatccaatgctactctgagcaatatctcctaagcttgagac 660  
 Db 601 tcatagcttcccgagaatccaatgctactctgagcaatatctcctaagcttgagac 660

Db 601 tcatagcttcccgagaatccaatgctactctgagcaatatctcctaagcttgagac 660  
 QY 661 tgaagatcgtatgattatatttcggatctcttagctgcgaagaaagcgatctcgag 720  
 Db 661 tgaagatcgtatgattatatttcggatctcttagctgcgaagaaagcgatctcgag 720  
 QY 721 tgcacagcttgcagatcgaagaatacacaacaaagccttctccgacacttagaa 780  
 Db 721 tgcacagcttgcagatcgaagaatacacaacaaagccttctccgacacttagaa 780  
 QY 781 ttgctcaagagtggtgctcctcccaagatcaagaagcttatctttagggaagc 840  
 Db 781 ttgctcaagagtggtgctcctcccaagatcaagaagcttatctttagggaagc 840  
 QY 841 taagatgctcagagctactacaataaagaacatgagagcctgactgtagatg 900  
 Db 841 taagatgctcagagctactacaataaagaacatgagagcctgactgtagatg 900  
 QY 901 caattagagagctcgaagcttcaattgcttgggaaagagagagccttccg 960  
 Db 901 caattagagagctcgaagcttcaattgcttgggaaagagagagccttccg 960  
 QY 961 gataaaagcaagcacttgagagcgctcgagcctgagcctgtagcttccg 1020  
 Db 961 gataaaagcaagcacttgagagcgctcgagcctgagcctgtagcttccg 1020  
 QY 1021 ctctgagatgagagctcgaagcttccgctcgagatctcctaaacactgaagagagc 1080  
 Db 1021 ctctgagatgagagctcgaagcttccgctcgagatctcctaaacactgaagagagc 1080  
 QY 1081 caagtgaatgagcttgaagctctcttagagtgtaggttgtagcaccctaaactgga 1140  
 Db 1081 caagtgaatgagcttgaagctctcttagagtgtaggttgtagcaccctaaactgga 1140  
 QY 1141 atacattacggaagcttgctcacaacacataataatgagactagctgagttctc 1200  
 Db 1141 atacattacggaagcttgctcacaacacataataatgagactagctgagttctc 1200  
 QY 1201 taaggggctacttcaaaaatctggaagcggtggaactatagcttccctaaagctccca 1260  
 Db 1201 taaggggctacttcaaaaatctggaagcggtggaactatagcttccctaaagctccca 1260  
 QY 1261 ggaaggggaagagtggtgctccacaacccgaggtctggaagagagatcttcttct 1320  
 Db 1261 ggaaggggaagagtggtgctccacaacccgaggtctggaagagagatcttcttct 1320  
 QY 1321 ctccgctacctaagaagcttaccctccctgatttataagcttggcgagctaga 1380  
 Db 1321 ctccgctacctaagaagcttaccctccctgatttataagcttggcgagctaga 1380  
 QY 1381 aactcagcttgcaactgagatcttctttaaagctacactcaacatcagaagcctt 1440  
 Db 1381 aactcagcttgcaactgagatcttctttaaagctacactcaacatcagaagcctt 1440  
 QY 1441 agactccttccaaagctggaagcttccctggaagacatcatccgagccttagaga 1500  
 Db 1441 agactccttccaaagctggaagcttccctggaagacatcatccgagccttagaga 1500  
 QY 1501 tcttgctatttataatctacaagaatctctgaaagaaacagttctcctcactgata 1560  
 Db 1501 tcttgctatttataatctacaagaatctctgaaagaaacagttctcctcactgata 1560  
 QY 1561 aaaaagactaattcagaaaccttcttcttggagcaggaacaaacaaagacccatcc 1620  
 Db 1561 aaaaagactaattcagaaaccttcttcttggagcaggaacaaacaaagacccatcc 1620  
 QY 1621 cagcatgacctatcaggttatacaggtacccagaaagcgtagagctcatgttgg 1680  
 Db 1621 cagcatgacctatcaggttatacaggtacccagaaagcgtagagctcatgttgg 1680  
 QY 1681 tattctagagacactagcaccctggaagcttccggaagatataccgttattatgata 1740  
 Db 1681 tattctagagacactagcaccctggaagcttccggaagatataccgttattatgata 1740



Db 1681 tatctagaagacactagaccctcgaagcttcctcgaagataccgttatgatatacaact 1740  
QY 1741 gatgacggaagagatgcacaataattccagctccttcagagctaccataaaattgt 1800  
Db 1741 gatgacggaagagatgcacaataattccagctccttcagagctaccataaaattgt 1800  
QY 1801 ggaagtaaccccaactacgtctatgaagcttgcttcttctatctagctccttgctc 1860  
Db 1801 ggaagtaaccccaactacgtctatgaagcttgcttcttctatctagctccttgctc 1860  
QY 1861 ttatgggttctcctcagctgagcttgcctcctcacaagagctgtgaa 1907  
Db 1861 ttatgggttctcctcagctgagcttgcctcctcacaagagctgtgaa 1907  
RESULT 2  
AAD03026/c  
ID AAD03026 standard; DNA: 36 BP.  
AC AAD03026;  
XX  
XX  
DT 13-JUN-2001 (first entry)  
XX  
DE Chlamydia pneumoniae OMP gene amplifying 3' PCR primer.  
XX  
KW Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;  
KM antibody; medicament; Chlamydia infection; PCR primer; ss.  
OS Chlamydia pneumoniae.  
XX  
PN WO200121804-A1.  
XX  
PD 29-MAR-2001.  
XX  
PF 15-SEP-2000; 2000WO-CA01088.  
XX  
PR 20-SEP-1999; 99US-0154652.  
XX  
PA (AVET ) AVENTIS PASTEUR LTD.  
XX  
PI Murdin AD, Oomen RP, Wang J, Dunn P;  
XX  
DR WPI: 2001-244939/25.  
XX  
PS Claim 32; Page 48; 82pp; English.  
XX  
CC The present sequence is a 3' PCR primer used for amplifying  
CC Chlamydia pneumoniae OMP (outer membrane protein) gene. This primer  
CC contains a Bam HI restriction site and the sequence encoding the  
CC C-terminal sequence of OMP. OMP is useful in the production of vaccines,  
CC antibodies and pharmaceutical compositions which are useful for treating  
CC or preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or  
CC C. pecorum) infections. They are also useful as diagnostic reagents for  
CC detecting Chlamydia infection which involves assaying a body fluid  
CC of a mammal to be tested for the components. The OMP vaccine is  
CC useful in the preparation of a medicament for preventing and/or  
CC treating Chlamydia infection. The primers derived from OMP gene are  
CC also useful for detecting and/or identifying Chlamydia in a biological  
CC material. OMP antibodies are also useful as reagents for purifying OMP  
CC from a biological sample which involves carrying out antibody-based  
CC affinity chromatography with the biological sample. OMP gene is  
CC also useful in gene therapy.  
XX  
SQ Sequence 36 BP; 8 A; 12 C; 7 G; 9 T; 0 other;

Query Match 1.3%; Score 24; DB 22; Length 36;  
Best Local Similarity 100.0%; Pred. No. 0.071;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1781 ggtctactcacaataatgtgag 1804  
Db 36 GGCTTACTCATTAATAATGTGAG 13  
RESULT 3  
AAD03025  
ID AAD03025 standard; DNA: 44 BP.  
XX  
AC AAD03025;  
XX  
DT 13-JUN-2001 (first entry)  
XX  
DE Chlamydia pneumoniae OMP gene amplifying 5' PCR primer.  
XX  
KW Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;  
KM antibody; medicament; Chlamydia infection; PCR primer; ss.  
OS Chlamydia pneumoniae.  
XX  
PN WO200121804-A1.  
XX  
PD 29-MAR-2001.  
XX  
PF 15-SEP-2000; 2000WO-CA01088.  
XX  
PR 20-SEP-1999; 99US-0154652.  
XX  
PA (AVET ) AVENTIS PASTEUR LTD.  
XX  
PI Murdin AD, Oomen RP, Wang J, Dunn P;  
XX  
DR WPI: 2001-244939/25.  
XX  
PS Claim 32; Page 48; 82pp; English.  
XX  
CC The present sequence is a 5' PCR primer used for amplifying  
CC Chlamydia pneumoniae OMP (outer membrane protein) gene. This primer  
CC contains a Not I restriction site, a ribosome binding site, an  
CC initiation codon and a sequence at the 5' end of the OMP coding  
CC sequence. OMP is useful in the production of vaccines, antibodies  
CC and pharmaceutical compositions which are useful for treating or  
CC preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or  
CC C. pecorum) infections. They are also useful as diagnostic reagents for  
CC detecting Chlamydia infection which involves assaying a body fluid  
CC of a mammal to be tested for the components. The OMP vaccine is  
CC useful in the preparation of a medicament for preventing and/or  
CC treating Chlamydia infection. The primers derived from OMP gene are  
CC also useful for detecting and/or identifying Chlamydia in a biological  
CC material. OMP antibodies are also useful as reagents for purifying OMP  
CC from a biological sample which involves carrying out antibody-based  
CC affinity chromatography with the biological sample. OMP gene is  
CC also useful in gene therapy.  
XX  
SQ Sequence 44 BP; 12 A; 14 C; 8 G; 10 T; 0 other;

Query Match 1.3%; Score 24; DB 22; Length 44;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 atgggaattcatctactctc 124  
Db 21 atgggaattcatctactctc 44



XX 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 PI Griffiths R;  
 XX  
 DR WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 PS Page 1677; Disclosure: 1912pp; English.  
 XX  
 CC AAX91991-X97517 represent PCR primers used to amplify open reading  
 CC frames and other nucleic acid sequences from the genome of  
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
 CC disease such as pneumonia and bronchitis and is thought to be a  
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
 CC containing C. pneumoniae nucleotides sequences can also be used as  
 CC immunogenic compositions, especially where the vector directs the  
 CC expression of a neutralising epitope of C. pneumoniae.  
 CC  
 SQ Sequence 20 BP; 3 A; 6 C; 3 G; 8 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1541 cgtctctcatgattatgc 1560  
 DB 1 cgtctctccatgattatgc 20

RESULT 7  
 ID AAX92063/C  
 XX AAX92063 standard; DNA; 20 BP.  
 AC AAX92063;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; PCR primer; ss.  
 XX  
 OS Synthetic.  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 PI Griffiths R;  
 XX  
 DR WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 PS Page 1482; Disclosure: 1912pp; English.

XX AAX91991-X97517 represent PCR primers used to amplify open reading  
 CC frames and other nucleic acid sequences from the genome of  
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
 CC disease such as pneumonia and bronchitis and is thought to be a  
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
 CC containing C. pneumoniae nucleotides sequences can also be used as  
 CC immunogenic compositions, especially where the vector directs the  
 CC expression of a neutralising epitope of C. pneumoniae.  
 CC  
 SQ Sequence 20 BP; 3 A; 7 C; 4 G; 6 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1288 ccgaggtcttgaagagcaga 1307  
 DB 20 CCGAGGTCTTGAAGAGCAGA 1

RESULT 8  
 ID AAX92065/C  
 XX AAX92065 standard; DNA; 20 BP.  
 AC AAX92065;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; PCR primer; ss.  
 XX  
 OS Synthetic.  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 PI Griffiths R;  
 XX  
 DR WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 PS Page 1482; Disclosure: 1912pp; English.  
 XX  
 CC AAX91991-X97517 represent PCR primers used to amplify open reading  
 CC frames and other nucleic acid sequences from the genome of  
 CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory  
 CC disease such as pneumonia and bronchitis and is thought to be a  
 CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent  
 CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded  
 CC by the open reading frames of the C. pneumoniae genome (see AAY34584-  
 CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors  
 CC containing C. pneumoniae nucleotides sequences can also be used as  
 CC immunogenic compositions, especially where the vector directs the  
 CC expression of a neutralising epitope of C. pneumoniae.  
 CC

5Q Sequence 20 BP; 4 A; 2 C; 10 G; 4 T; 0 other;

## Query Match

Best Local Similarity 100.0%; Score 20; DB 20; Length 20;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 319 catcacctccatccatccg 338  
|||||

DB 20 CATCCACTCCTCAGATCCGC 1

## RESULT 9

AAx92068/c  
ID AAX92068 standard; DNA; 20 BP.

AC AAX92068;

DT 13-SEP-1999 (first entry)

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

OS Synthetic.

OS Chlamydia pneumoniae.

PN MO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GEST ) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AAx92070/c  
ID AAX92070 standard; DNA; 20 BP.

AC AAX92070;

DT 13-SEP-1999 (first entry)

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

OS Synthetic.

OS Chlamydia pneumoniae.

PN MO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GEST ) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AAx92070/c  
ID AAX92070 standard; DNA; 20 BP.

AC AAX92070;

DT 13-SEP-1999 (first entry)

DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope; PCR primer; ss.

OS Synthetic.

OS Chlamydia pneumoniae.

PN MO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GEST ) GENSET.

PI Griffiths R;

DR WPI; 1999-357842/30.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 20 BP; 4 A; 4 C; 8 G; 4 T; 0 other;

Query Match 1.0%; Score 20; DB 20; Length 20;

Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1272 ggttgcttcacacccga 1291  
|||||

DB 20 GGTGCTCTCCACACCGA 1

## RESULT 11

AA216609  
ID AA216609 standard; cDNA; 800 BP.

AC AA216609;

DT 12-OCF-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:4079.

KW Human; gene; gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.  
 XX  
 XX MO9938972-AZ.  
 XX  
 XX 05-AUG-1999.  
 PD  
 XX  
 XX 28-JAN-1999; 99WO-US01619.  
 PF  
 XX 03-APR-1998; 98US-0080666.  
 PR 28-JAN-1998; 98US-0072910.  
 PR 24-FEB-1998; 98US-0075954.  
 PR 31-MAR-1998; 98US-0080114.  
 PR 03-APR-1998; 98US-0080515.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Ctkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
 PI Escobedo J, Garcia A, Garcia V, Glese K, Inlis MA;  
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
 PI Stache-Crain B, Suduth-Klinger J, Williams LT;  
 DR WPI; 1999-494092/41.  
 XX  
 XX Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 XX  
 XX Claim 1; Page 1934; 2479pp; English.  
 XX  
 XX The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AA212532 to AA217779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein); and to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 CC  
 XX Sequence 800 BP; 199 A; 166 C; 202 G; 211 T; 22 other;  
 XX  
 XX  
 XX Query Match 1.0%; Score 19; DB 20; Length 800;  
 XX Best Local Similarity 100.0%; Pred. No. 29;  
 XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 351 gcacatcattgagcagc 369  
 Db 500 gcacatcattgagcagc 518  
 AAD07852 standard; cDNA; 1210 BP.  
 ID AAD07852; 1210 BP.  
 XX  
 AC AAD07852;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Human secreted protein-encoding gene 6 cDNA clone HEITP82, SEQ ID NO:54.

XX  
 KW Human: secreted protein; proliferative disorder; cancer; tumour; asthma;  
 KW foetal abnormality; developmental abnormality; hematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;  
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;  
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;  
 KW angigenic disorder; kidney disorder; gastrointestinal disorder; allergy;  
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;  
 KW cell culture; chemotaxis; vulnery; binding partner identification;  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX  
 XX Key Location/Qualifiers  
 FT complement (770..1129)  
 FT CDS  
 FT /tag- a  
 FT /product= "Human secreted protein precursor"  
 FT /note= "CDS does not include start codon"  
 FT /partial  
 FT sig-peptide complement (1127..1129)  
 FT /tag- b  
 FT mat-peptide complement (773..1126)  
 FT /tag- c  
 FT /product= "Mature human secreted protein"  
 PN  
 XX  
 XX WO200132675-A1.  
 PD  
 XX  
 XX 10-MAY-2001.  
 XX  
 XX 25-OCT-2000; 2000WO-US29363.  
 PF  
 XX  
 XX 29-OCT-1999; 99US-0162239.  
 PR 30-JUN-2000; 2000US-0215139.  
 PR  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ruben SM, Komatsoulis GA, Wei P, Baker KP, Young PE;  
 PI WPI; 2001-328772/34.  
 DR P-PSDB; AAE03470.  
 DR  
 XX  
 XX Thirty two human secreted proteins, useful for treating cancers,  
 PT hyperproliferative disorders, inflammatory disorders, neurological  
 PT disorders, autoimmune diseases and cardiovascular disorders -  
 XX  
 XX Claim 1; Page 438; 576pp; English.  
 PS  
 XX  
 XX AAD07809-AAD07907 represent cDNAs corresponding to 32 human secreted  
 CC protein genes, and AAE03427-AAE03523 represent the proteins they encode.  
 CC AAE03524-AAE03537 represent human secreted protein fragments or variants.  
 CC  
 CC The secreted proteins and their genes are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 32 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC hematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angioleptic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. A6564197-A6594564 represent novel human diagnostic coding sequences of the invention.

DR	MP1: 2001-457740/49.
DR	P-PSDB: ABB11957.
XX	
PT	Human proteins and DNA encoding sequences useful for preventing,
PT	treating or ameliorating a medical condition in a mammalian subject
XX	e.g. arthritis and cancer -
PS	Claim 1: Page 835: 1963pp: English.
XX	
XX	Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC	sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC	invention also relates to vectors and recombinant host cells comprising a
CC	nucleotide of the invention, methods of producing the novel polypeptides
CC	antibodies against the polypeptides, methods of detecting the nucleotides
CC	or polypeptides in a sample, and methods of identifying compounds which
CC	bind to polypeptides of the invention. Although novel, many of the
CC	polypeptides of the invention have homology to known proteins, thereby
CC	giving an insight into their probable biological activities, and hence

potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haemopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.

Sequence 1958 BP; 499 A; 418 C; 424 G; 617 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1958;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 351 gcaccatcattggagcagg 369  
|||||  
DB 226 gcaccatcattggagcagg 244

RESULT 15

AAH98573 standard; cDNA; 1958 BP.

AAH98573;

12-OCT-2001 (first entry)

Human EST-derived coding sequence SEQ ID NO: 430.

Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

diagnostics; forensic test; gene mapping; genetic disorder;

biodiversity; gene therapy; nutrition; ss.

Homo sapiens.

WO200154477-A2.

02-AUG-2001.

25-JAN-2001; 2001WO-US02687.

25-JAN-2000; 2000US-0491404.

17-JUL-2000; 2000US-0617746.

03-AUG-2000; 2000US-0631451.

15-SEP-2000; 2000US-0663870.

(HYSE-) HYSEQ INC.

Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,

Cao Y, Drmanac RA, Zhang J, Werhman T;

WP1; 2001-476164/51.  
DR P-PSDB; AAM23914.

PT isolated polypeptide for treatment of diseases, diagnostics, raising

antibodies and research use -

Claim 1; Page 484; 1275pp; English.

The present invention provides the protein and coding sequences of novel proteins from a variety of organisms, including human, dog, cat, horse, cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin and tomato. These were derived from expressed sequence tags (ESTs) from the organism of interest. They can be used in diagnostics, forensics, gene mapping, identification of mutations, to assess biodiversity and for nutritional purposes. The present sequence is a cDNA of the invention.

Sequence 1958 BP; 499 A; 418 C; 424 G; 617 T; 0 other;

Query Match 1.0%; Score 19; DB 22; Length 1958;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 351 gcaccatcattggagcagg 369  
|||||  
DB 226 gcaccatcattggagcagg 244

Search completed: August 20, 2002, 03:16:11  
Job time: 4616 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:20:55 ; Search time 42.94 Seconds  
(without alignments)  
1271.047 Million cell updates/sec

Title: US-09-662-812-2  
Perfect score: 2833  
Sequence: 1 MGLFHLTFGLLCSLPSL.....TEGDAKNFVLGLIKIVE 568

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR-71:\*  
1: pirl:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2833	100.0	568	2 C72129	probable outer mem
2	2833	100.0	568	2 G86493	probable leader pe
3	1456	51.4	566	2 A81682	conserved hypothe
4	1421	50.2	566	2 G71525	probable outer mem
5	203	7.2	1547	2 A12043	hypothetical prote
6	164	5.8	1119	2 AB2239	hypothetical prote
7	153	5.4	1348	2 AG2558	hypothetical prote
8	148.5	5.2	2297	2 AB2494	hypothetical prote
9	145	5.1	1010	2 AH2553	hypothetical prote
10	145	5.1	1381	2 AF2010	regulatory protein
11	134.5	4.7	730	2 S28294	hypothetical prote
12	134.5	4.7	826	2 C88550	hypothetical prote
13	134.5	4.7	1238	1 S17944	protein ZC84.3 (im
14	134.5	4.7	1238	1 S17946	virulence sensor p
15	134	4.7	860	2 T37768	virulence sensor p
16	133	4.7	321	2 F83358	probable vacuolar
17	133	4.7	1069	2 AF1930	hypothetical prote
18	133	4.7	1138	2 G71554	hypothetical prote
19	131.5	4.6	2469	2 H36812	probable transmem
20	131	4.6	398	2 T44927	hypothetical prote
21	131	4.6	398	2 AD1882	hypothetical prote
22	129.5	4.6	2337	2 T40577	hypothetical prote
23	128	4.5	1189	2 A54817	probable phosphati
24	127	4.5	1238	1 A40185	ATPase Scitl, chrom
25	126.5	4.5	1292	2 D84727	virulence protein
26	126	4.4	631	2 JCA298	probable RAD50 DNA
27	126	4.4	789	2 T38423	hyaluronan recepto
28	126	4.4	851	2 AG2469	hypothetical prote
29	125	4.4	1805	1 A64224	hypothetical prote

30	124.5	4.4	526	2 B81679	conserved hypothe
31	124	4.4	1275	2 T49362	hypothetical prote
32	124	4.4	1708	2 AE1866	WD-40 repeat prote
33	123.5	4.4	894	2 T13029	beta-adaptin homol
34	122	4.3	719	2 T05384	beta-adaptin homol
35	121.5	4.3	556	2 B46024	neurofilament-L su
36	121.5	4.3	866	2 A11880	aminopeptidase (im
37	121.5	4.3	2555	2 C69681	peptide synthetase
38	121	4.3	566	2 G75457	tetratricopeptide
39	121	4.3	1603	2 VUKWS	vitellinogen vit-5
40	121	4.3	1603	2 F89497	protein vit-5 (imp
41	120.5	4.3	430	2 F75053	hypothetical prote
42	120.5	4.3	937	2 A35553	beta-adaptin - hum
43	120.5	4.3	937	2 C35553	beta-adaptin - rat
44	120	4.2	922	2 G83109	probable two-compo
45	119.5	4.2	1466	2 G84516	probable retroelem

ALIGNMENTS

RESULT	1	
C72129	1	
Probable outer membrane leader peptide (omp) CP0021 - Chlamydia pneumoniae (str		
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae		
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000		
C:Accession: C72129, B81542		
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,		
Nature Genet. 21, 385-389, 1999		
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.		
A:Reference number: A72000; MUID:99206606		
A:Accession: C72129		
A:Molecule type: DNA		
A:Residues: 1-568 <ARN>		
A:Cross-references: GB:AE001587; GB:AE001363; MID:94376271; PIDN:AD18174.1; PID:943		
A:Experimental source: strain CML029		
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick		
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salz		
Nucleic Acids Res. 28, 1397-1406, 2000		
A:Title: Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR3		
A:Reference number: A81500; MUID:20150235		
A:Accession: B81542		
A:Molecule type: DNA		
A:Residues: 1-568 <REA>		
A:Cross-references: GB:AE002234; GB:AE002161; MID:97189667; PIDN:AF38558.1; PID:971		
A:Experimental source: strain AR39, HL cells		
C:Genetics:		
A:Gene: CP0021; CP0755		
Query Match		
Best Local Similarity 100.0%; Score 2833; DB 2; Length 568;		
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	MGLFHLTFGLLCSLPSLVAKFPESVGHKILVISTOSTOQALATYLEALDAYGDHDF 60
DB	1	MGLFHLTFGLLCSLPSLVAKFPESVGHKILVISTOSTOQALATYLEALDAYGDHDF 60
QY	61	VLRKIGEDYLKQSHSSDPQTRKSTIIGAGLGSSEALDVLSQAMETADPIQQLVLVSAY 120
DB	61	VLRKIGEDYLKQSHSSDPQTRKSTIIGAGLGSSEALDVLSQAMETADPIQQLVLVSAY 120
QY	121	SGHLGKTSDDLFLFKALASPYVIRLEAAYRLANKTKVVIDHLHSFHKRLPEEIOCLSA 180
DB	121	SGHLGKTSDDLFLFKALASPYVIRLEAAYRLANKTKVVIDHLHSFHKRLPEEIOCLSA 180
QY	181	IFLRLEPESDAYIRDLAAKSAIRSATALQIGGYOQKRFPLTDLNLTSSAPDQDAI 240
DB	181	IFLRLEPESDAYIRDLAAKSAIRSATALQIGGYOQKRFPLTDLNLTSSAPDQDAI 240
QY	241	LYALGKLKDGQSYNIKKQLOKPDVDVTLAAQAALIAKGEDALPVYKKALERRPAL 300
DB	241	LYALGKLKDGQSYNIKKQLOKPDVDVTLAAQAALIAKGEDALPVYKKALERRPAL 300

Qy	301	YARHLBEICIPIALPIFLKTKNSEKLNVALLELGCOTPKLEYITERLVOPHYNE	360
Db	301	YALRLRBEICIPIALPIFLKTKNSEKLNVALLELGCOTPKLEYITERLVOPHYNE	360
Qy	361	TLALSFSGRILQWKKRNIIVPDDPDERRELTSTRGLEQILTELFRTPKEAVLPCY	420
Db	361	TLALSFSGRILQWKKRNIIVPDDPDERRELTSTRGLEQILTELFRTPKEAVLPCY	420
Qy	421	KLASQKQLATTALISFLSHTSHOEADLLFOAKLPGEPITRAYADLAIYNLTKPEKK	480
Db	421	KLASQKQLATTALISFLSHTSHOEADLLFOAKLPGEPITRAYADLAIYNLTKPEKK	480
Qy	481	RLSLHDYAKKLIQLELLEVDENORPHEPSMYLROVYTPESRTKLMDIETLATSSED	540
Db	481	RLSLHDYAKKLIQLELLEVDENORPHEPSMYLROVYTPESRTKLMDIETLATSSED	540
Qy	541	IRLLIQIMTEGDANKPVLGGLIKLIVE	568
Db	541	IRLLIQIMTEGDANKPVLGGLIKLIVE	568

Query	Subject	Score	Length	Mismatches	Conservative	Indels	Gaps
1	MGELHLLTFGLGLCSLPSLVAKFPESGKILITISQSTQCALATYLEALDAYGDHDF	100.0%	568	0	0	0	0
2	1 MGLHLLTFGLGLCSLPSLVAKFPESGKILITISQSTQCALATYLEALDAYGDHDF	100.0%	568	0	0	0	0
3	61 VLRKIGEDYLKQSHSSDPQTRKSTIIIGAGIAGSSSEALDVLQAMETADPLQQLLVSAV	100.0%	568	0	0	0	0
4	61 VLRKIGEDYLKQSHSSDPQTRKSTIIIGAGIAGSSSEALDVLQAMETADPLQQLLVSAV	100.0%	568	0	0	0	0
5	121 SGHLGKTSDDLKFKALASPPVIRLEAAVRLANKNTRVIDHLHSFTIHKLPPEIIOCSAA	100.0%	568	0	0	0	0
6	121 SGHLGKTSDDLKFKALASPPVIRLEAAVRLANKNTRVIDHLHSFTIHKLPPEIIOCSAA	100.0%	568	0	0	0	0
7	181 IFLRLTEESAAVYRDLIAKKSARSATALQIGEYQKRLPLRLNLLSASQDDEAI	100.0%	568	0	0	0	0
8	181 IFLRLTEESAAVYRDLIAKKSARSATALQIGEYQKRLPLRLNLLSASQDDEAI	100.0%	568	0	0	0	0
9	241 LVALGKTKDQSYNNIKKQLOKPDVDTLAAQAALILAKKEEDLPIYKQALBERPRAL	100.0%	568	0	0	0	0
10	241 LVALGKTKDQSYNNIKKQLOKPDVDTLAAQAALILAKKEEDLPIYKQALBERPRAL	100.0%	568	0	0	0	0
11	301 YALRHLPSEIGIPALPIFLTKNSEAKLNVALLLELGDTPKILEYITERLVQPHYNE	100.0%	568	0	0	0	0
12	301 YALRHLPSEIGIPALPIFLTKNSEAKLNVALLLELGDTPKILEYITERLVQPHYNE	100.0%	568	0	0	0	0
13	361 TLALSPKSGKTLNWKVNNIIVDDQPERERLLSTTGLEBQILTFELRPLKRAYLYCII	100.0%	568	0	0	0	0
14	361 TLALSPKSGKTLNWKVNNIIVDDQPERERLLSTTGLEBQILTFELRPLKRAYLYCII	100.0%	568	0	0	0	0

QY	421	KLASQKQTALTTAISLSHTSHOEADLFLQAAKLEGEPIIRAYADLAIYNLTKEPK	480
DB	421	KLASQKQTALTTAISLSHTSHOEADLFLQAAKLEGEPIIRAYADLAIYNLTKEPK	480
QY	481	RSLSHDYAKKLIQETLLEVDENORPHDSMPYLRYQVPESEKTKIMDILETATSSSED	540
DB	481	RSLSHDYAKKLIQETLLEVDENORPHDSMPYLRYQVPESEKTKIMDILETATSSSED	540
QY	541	IRLLIQLMTBEGDANKFVLAALLIKIYE	568
DB	541	IRLLIQLMTBEGDANKFVLAALLIKIYE	568

3

rsult

1682

conserved hypothetical protein TC0629 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C:Accession: A81682

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gynn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; et al. (2000) Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255

A:Accession: A81682

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-566 <TEET>

A:Cross-references: GB:AE002331; GB:AE002160; MID:g7190662; PIDN:AAF39458.1; PID:g719

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0629

Query Match 51.4%; Score 1456; DB 2; Length 566;

Best Local Similarity 53.5%; Pred. No. 1,Je-80;

Matches 304; Conservative 92; Mismatches 170; Indels 2; Gaps 1;

QY 1 MGLPHLTLFGLLSLPSLVAKPFPESVGHKILYITSTQQAATATYALADAGDHPDF 60

Db 1 MGLSRLLIFGL--SLPLASCDPFPSPVSKILFLCKSKTPQALESTYLENSTTYQCHNFS 58

QY 61 VLRKIGEDYKQSTHSSDPOTRKSTTIGAGIAGSSSEALDVLISQAMETADPLQOLLVLSAV 120

Db 59 ILRLIASYLDQOSTFSSDAYVRKSAIIGAGISGSSSETLDLSESIETQDYLEQHLINAA 118

QY 121 SGHIGKTSDDLFFALASPYVIRLEAAVIRLANKKRVKIDHSHFHKLPPEIQCISAA 160

Db 119 GNOIGKTSDDLFFGLAPRPIITLEAAVIRLCKKNSKVSVDYLSFHQLPPEIQLAAT 178

QY 181 IFLRLEESDAYIRDLLAAKKSARSATALQIGEVQAKRFLPTRLNLTLSASPODEAT 240

Db 179 IFLQLETEADAYVIRLLSSPNSLTRNMYAVLIGEVQGRFLPTRLSILTSAPPLDEGS 238

QY 241 LYLAKGKLDQSGYYNIRKQLOKRPVDYTLAAQAQLALGKEEDALPVIRKQALEERPAL 300

Db 239 LYLAKGKLEDDSSYPKIRALSKSNPEVALAAQQLTLFGKEDALPLITTFCCQOEPRAL 298

QY 301 YALRHLESEIGIPALPIFLFKTKNSEAKLNVALLALEGCDPRLLEYITERLVOPHYNE 360

Db 299 YTSRFLSLKEGDELLPIFPCKAIKEIKLNAALALVHLSGVNHLVLSYLFELFNKILHR 358

QY 361 TLALSSKSGRTLQNMKRVNITIVDDPOERERLLSTTRGLEEQILTFELRLPKRAYLPCYI 420

Db 359 IFLPETHIGATQFWKCTCALPLSLPPEEKARALAMRAAEDTILSSSLKLPNNAYLPYLE 418

QY 421 KLASOKTQATATASLSTSHOEDLILFOAKILGPEITIRAYVDLAIYNTLTKPPEK 480

Db 419 RILTSQTPPLAAKALIALSVTAHPQALSLVSKAALPLGDEPTIRAVANLALYTTQDPCK 478

QY 481 RSLHDYAKKLIQETLLLVDTENORPHPSMPLYLYQVTPESRTLMIMDILETTATSKSSD 540

```

Db 479 ALXQVAEOLIGDITLETFDEBNPLPSHSSYLKQVSPFTRSQLMTILETVSSKTTED 538
QY 541 IRLLIQMTESGDAKNPVLGLIKIYE 568
   ||: : ||: || : ||: ||: ||: ||
Db 539 IRVFLSLMKTHYKNIPILSGLMRIYE 566

```

#### RESULT 4

Probable outer membrane leader peptide (omp) CT350 - *Chlamydia trachomatis* (serotype D,  
 C:Species: *Chlamydia trachomatis*  
 C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 04-Feb-2000  
 C:Accession: G71525  
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
 Science 282, 754-759, 1998  
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trach*  
 A:Reference number: A71570; MUID:99000809  
 A:Accession: G71525  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-566 <ARN>  
 A:Cross-references: GB:AE001308; GB:AE001273; NID:g3328766; PIDN:AAC67945.1; PID:g332877  
 A:Experimental source: serotype D, strain UW-3/Cx  
 C:Genetics:  
 A:Gene: CT350

Query Match	50.2%	Score 1421	DB 2	Length 566
Best Local Similarity	53.0%	Prod. No. 1	56-78	
Matches 301; Conservative	88;	Mismatches 177;	Indels 2;	Gaps 1

hypothetical protein alr1903 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: A12043  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matsubae, A.; Iqbal  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: A12043  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1547 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA073602.1; PID:g17130993; GSPDB:GN00179  
C:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr1903

Query Match	7.2%	Score 203;	DB 2;	Length 1547;
Best Local Similarity	21.3%	Pred. No. 0.00037;		
Matches 114; Conservative	105;	Mismatches 239;	Indels 76;	Gaps 16;

## RESULT 6

**hypothetical protein ali13465** [imported] *Anabaena* sp. (strain PCC 7120)  
**C**:Species: *Anabaena* sp.  
**A**:Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp., strain PCC 7120  
**C**:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
**C**:Accession: A82293  
**R**:Kakeko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irlig,  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat,  
 DNA Res. 8, 205-213, 2001  
**A**:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AB2239  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1119 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA875164.1; PID:g17132598; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all13465

Query Match 5.8%; Score 164; DB 2; Length 1119;  
 Best Local Similarity 23.2%; Pred. No. 0.054;  
 Matches 95; Conservative 65; Mismatches 180; Indels 70; Gaps 13;

95 SEALVLSOMETADPDLQOLVLSAVSGHUG-----KTSDDLFLKALASPY 140  
 DB 699 SOVFOILCSLNTETAFOTOLQLKREPOIAQVRFQYCAALGKKAADRLLELLKHS 758  
 QY 141 PVIRLEAAVRLANKNTKYIDHLHSFIHKLPPEIQCLSAIFL-RLTEVESDAYIRDLA 199  
 DB 759 SNVRSSADALGNITENGIPGLEFL-KDPEFNWFSFTNALGNIGTAAIPGLELLK 817  
 QY 200 AKKSARSATALOIGEYOQKRFPLRLNLTSASPODEALILYALGKL-----247  
 DB 818 HSESNVRSSAAEALGKIGAESAIPLLELLKHSSESNVRSSAAEALGKIGAESAIIDLEL 877  
 QY 248 -KDGGSY-----NIKLOKPDVDTLAAOALIALGKE--EDAL 285  
 DB 878 FKDSISIVRSSAINALSKIGAEIAPDLKLKLDSESVRSSAAEALGQIGAEIAPDL 937  
 QY 286 PVIKQALEERPRALYRLHPSEIGIPIALPIPLKTNSEKLNVALLELG--CQTP 343  
 DB 938 KLINDSEYVFKAAVVALGQIGAEIAPDLKL-LKNSNFTVRCGALALDIGNAAIP 996  
 QY 344 KLEVTERLVQPHYNETLALSFSKRTLQNMKRVNIIVPODERERLSTTGL-EEQ 402  
 DB 997 DLKLKLDSEYVRSAAVAL-----GQI-----GAEIAP-----DLKLINDSEYVSS 1043  
 QY 403 ILTFLRLPKRAVLPCTIKLLASOKTQATTAISFL---SHFSHOEALDL 449  
 DB 1044 AAVAGQIGAEATIPDLKLKLDSEYVRSAAEALGKIDERTHSVIVNL 1093

RESULT 7  
 AG2558  
 hypothetical protein all18062 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7120  
 C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: AG2558  
 R:Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi  
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
 DNA Res. 8: 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AG2558  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1348 <KUR>  
 A:Cross-references: GB:AP003603; PIDN:BA877392.1; PID:g17134835; GSPDB:GN00182  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all18062  
 A:Genome: plasmid

Query Match 5.4%; Score 153; DB 2; Length 1348;  
 Best Local Similarity 22.8%; Pred. No. 0.32;  
 Matches 163; Conservative 83; Mismatches 264; Indels 206; Gaps 31;  
 9 FGLLCSLPISLVAKPESEVGHKILYISTOSTQOALATYALDAYGDHDFVLRKIGED 68

DB 340 VALIIASLN-SLAANLPAOLLIALVQKNVTPQOGLAVYLOSSNPE-----NKA 387  
 QY 69 YLKQSHSSDPOTRKSTIIGAGSSSEALD-----VLSOMETADP 110  
 DB 388 YLLTQVNLPLPNLKLGLKALAAAREIODESNRAALSLAPLLPELSEALAAAE 447  
 QY 111 LQ-QLIVLSAVSGHGLKTSDDLFLKALA-----SPYVI-----RLBA 147  
 DB 448 IOAEIERAYALSLTPPLPELPSKALATAREIODESNVRSLSSLPPLPELPEALAA 507  
 QY 148 AYRIAN-LKNTKYIDHLHSFIHK-LPEIQCLSAIFLRLTEVESDAYIRDLA-----A 200  
 DB 508 AREIODEYKRAKALSLAPLPELPEAL- -AAVREIODEEGRAYLSLAPLPEL 562  
 QY 201 KKSARSATALOIGEYOQKRFPLRLNLTSASP-----ODEALYVALGKLDQ 251  
 DB 563 LPEALVAAKEFG-DEYRRAKALSLAPLPELPEALATREIODEEGAYLS-----615  
 QY 252 STYNIKKQOKPDVDTLAAOALIALGKEEDAL-----PVIKQALEERPRALYRLHP 307  
 DB 616 ---NLAPLLPELSEALAAAREFODESNRAALSLAPLPAELSE--ALVAAAREFO 669  
 QY 308 SEIGIPIAL-----PIFLKTNSEA-----KLNVALLELGODPKLL--EYTER 352  
 DB 670 SEYIRAKALSLAPLPELSEALAAAREFODESNRAALSLAPLPELPEALAAAR 729  
 QY 353 LVQPHYNETLALS-----FSKRTLQNMK-RVNIIVPOD-- 385  
 DB 720 EIOSEYDRALASRLAPLPELSEALAAARQIODEYRAKLTSSWGKLPPELLPALA 789  
 QY 386 -----POERERLSTTGLLEBOILTFRLPKRAVLPCTIKLLASOKTQATTAISFLSH 440  
 DB 790 AAREIODEYDRALASR-----LALL--LPELSEALAAARQIODEYRALALSLAP 841  
 QY 441 TSHOALDLLEQAA-----KLGPPIRAYADALATYINLKDEKRS 482  
 DB 842 LLPPELLSEVLAAREIONEYERANTLSSLPKLPPELLSEYLA--AAREIONEYRANT 899  
 QY 483 LHDYAKKLIQETLFLV-----DTENQRPHPMPYLRQVTPESRTKLMIDL--ETLATS 535  
 DB 900 LSSLPLNKLPELASKVLAAREIONE-----EYRALSLSLALLPKLPPELLAA 950  
 QY 536 K--SEEDR-----LLIQMTEG-----DAKNRPVYAGLLIKIV 567  
 DB 951 KEIOSEYDRALASRLAPLPELSEALAAAREIODEYRAKALSLAPLPEL 1006

RESULT 8  
 AB2494  
 hypothetical protein all17130 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7  
 C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: AB2494  
 R:Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriq  
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, A.  
 DNA Res. 8: 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AB2494  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2297 <KUR>  
 A:Cross-references: GB:BA000020; PIDN:BA878214.1; PID:g17135668; GSPDB:GN00180  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all17130  
 A:Genome: plasmid

Query Match 5.2%; Score 148.5; DB 2; Length 2297;  
 Best Local Similarity 22.2%; Pred. No. 1.3;  
 Matches 147; Conservative 89; Mismatches 205; Indels 221; Gaps 34;

```

OY 37 TOSTQALATYLBALDAYGDHDFVLRKIGEDYLKQSHSSDPQTRKSTIGAGLASSSE 96
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 334 TEVFOQALFALPKRIE-NGVNIFSI-----IAPQLTEFEQLLEKT----- 373
OY 97 ALDVLQAMETADPDLQQLLVLSAVSGHLGTSDDLFLKALASPPVRLAAVRLANLKN 136
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 374 -LEIVNN-LERGDYQREALV-AIAPYLPKSEPLLQQA-----LEIAVKIAN--- 417
OY 157 TKVIDHLSFIHLKP--EEIOGSLAAFIKRL-----TEESDAYITDILAARK 202
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 418 -EEDRKALAAIAPHLKKEQC-NEVLYLLEERENIWPVIGSVSTLVFIASYLSKSQ 475
OY 203 S-----AIRSATALQIGEYOQKRFPLRLNLTLSASPO-DOEALVYALGKLKDGQSYVNI 256
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 476 SKLLQKAFNIVQNLFEFGSYQALALVALAPHLSKLEPQLLOQALAKIAI-----NI 525
OY 257 KKOLOKPDVUTLAAQAALALGKEEDALPVYIKQALE-----RPRALYA- 302
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 526 KE-----GASKVVALFAVYIPHLPOSKQLLEKAFELIQTIEYDNASSRALVAV 573
OY 303 LRHLPS-----ELGIPIALPIFLKTKNSEAKLNVALLLEIGDPPKLEIYTERLYOP 356
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 574 IPIHLSPESHLDKSLTEIKLVSTYDSAKAQAALVAYA-----POLKRF-NPSLLQ- 623
OY 357 HYNETLALSFSKGTLLQNMKRVNIIVPODERERLSTTGLLEE--QITLFLFLPKKE 413
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 624 -----KALEIAKKIN-----SDEQKDEALAAIVQLSSESPELLEQYLKLAQN 666
OY 414 AYLPCLYK-----LLASOKTO-----LATTAISFLSHTS-----HOEALDDL 450
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 667 TYHSHKRIALIPYLRQOQTEFLQKYSATASLASAPCLPEPKRSEVLQEALEKML 726
OY 451 FOAKLPGEPITRAYADLAIYNTLKOPEKRSLLHDYAKKLIQET-----LLPQDT----- 500
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 727 LDANY-----SDLLRANEIKIIVPLSLKTEQEKYEIAKTKNDPFGCAELIAEVATHLSE 783
OY 501 -----ENORPHPS-----MPYLKQVTPESRTE-----LM 525
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 784 FDOYKIFELIERLKAIEENDSSKAKALAVIPIYLS-KSSPESIDKAFETAENQYOSQC 842
OY 526 LDILETLATS-KSESDIRLLIQLMTEG-----DAKNFPVLAG-----LLIKI 566
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 843 FDDLVTLATLHKEREICIKLEQALAKADIDSEYQQAQDPASITLSQPKSGSGIPLLYKI 902
OY 567 VE 568
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 903 LE 904

```

RESULT 9  
 AH2553  
 hypothetical protein all8023 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120  
 C:Species: Anabaena sp.  
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C:Accession: AH2553  
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AH2553  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 11010 <KUR>  
 A:Cross-references: GB:AP003603; PIDN:BAW77353.1; PID:g17134796; GSPDB:GN00182  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all8023  
 A:Genome: plasmid

```

Query Match 5.1%; Score 145; DB 2; Length 1010;
Best local similarity 22.3%; Pred. No. 0.66;
Matches 136; Conservative 84; Mismatches 210; Indels 180; Gaps 30;
OY 18 ISLVAKPEESVGKILYISTOSTQOALATYLEALDA--GDHDFVLRKIGEDYLKQSH 75
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 81 IALADKLPPDL-----LSEALTSAREIQDEYLCADALIALAKLPDLLEAL 129
OY 76 SS-----DPQTRKSTIGAGLASSSEALDVLSQAMETADPQQLLVLSAVSGHLGTSDDL 131
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 130 TAREIQDEYFRTSTLLE-----AEKLPVSYLEALAAAREIQ-----DE 168
OY 132 LFK-----ALASPPVRLAAVRLANLKNKY-IDHLSFIHLKPEEI--QCISAAIFL 183
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 169 YFRASTLILAEKLPVLSKSAALAAAREIQDEYFRADALRELQKLPDLLEALAAVREI 228
OY 184 RLEESDAYITDILAARKKSAISATLQIGEYOQKRF-LPTLRNLTLSASPOQEAELY 242
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 229 QPEYLRADALI--ALVKLPVSYLEALAAAREIQDEYLAADALRELQKLPDLLEAL 286
OY 243 ALGKLKDGQSYVNIKKOLOK---PD-VDVTLAAO-----ALIALGKE-----E 282
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 287 AATEIRGCGYHTNPLRELAEKLPDLLEALAAAREIQDESNRAHALRELAEKLPDLS 346
OY 283 DALPVIRK-----QALEER-----PRALVALRHLPSERG----- 311
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 347 EALTATREIQSEYHRASTLALAQKLPDLLEALAAAREIQDESNASTLRELAEKLP 406
OY 312 -IPIALPIFLKTKNSEAKLNVALLLE-IGCDIPKLEIYTERLYOVHYETLALSRSKG 369
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 407 VLPEALAAVRRIRKSNRAVALALAEKLPVSYLEALAAATE--IEPEYHRA----- 456
OY 370 RTLONMKRVNIIVQDERERLSTTGLLEQILFTFLPKPRAVYLPCLYKLLASOKTQ 429
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 457 -----STURELAE-----KLPPD-----LLSE--- 473
OY 430 LATTAISFLSHTSHOEALDLFOAKLPGEPITRAYADLAIYNTLKOPEKRSLLHDYAKK 489
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 474 -ALTAISEIQPKSNR-ADALIALAEKLPDLLEALAA--AIRIQDESNRAHALIALAEK 529
OY 490 ----LQETLLE----VTEORPHPSMPYLRQVTPESRTEKMLDIL-ETLANSK----S 537
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 530 LPPDLLEALAAAREIQDESNRAHALIALAQ-----KLPPDLLEALAAAREIQSK 580
OY 538 SEDIRLLIOL 547
   1: ||||| 1: ||||| 1: ||||| 1: |||||
Db 581 SNRVHALIAL 590

```

```

RESULT 10
AF2010
regulatory protein all1636 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AF2010
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2010
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 11381 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW78002.1; PID:g17135456; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1636

```

Query Match 5.1%; Score 145; DB 2; Length 1381;  
 Best local similarity 20.7%; Pred. No. 1;

Matches 122; Conservative 87; Mismatches 190; Indels 190; Gaps 23;

```

QY 20 LVAKPEESGHIKITYISTOSTOQALATYIEALDAYGDHPFV--LRKIEDVYKOSIHSS 77
Db 688 LVDKLEPEL-----PEALAAAREIENSYNRRDALISLANKLPPELLPEAVAAA 735
QY 78 ---DPQTRKSTIIGA-----GLASSEALDVLSSQA---METADPLOOL----- 115
Db 736 REIOEANNAEVLGSLDKLPPELLPEALAAAREIEFESSRATYILSLADKLPPELLPEALA 795
QY 116 -----VLANSVGH-----GKTSDDLKFA-----LASPPYVIRLE 146
Db 796 ATREIONGYNRADILSLADKLPPELLSEALAVAREIEDEKRYAREALSLADKLPPELLPE 855
QY 147 AAYRLANKNTKV-IDHHSFTHKLPPEIQCSAIFLR-LTESDAYIRBLAAK--- 201
Db 856 ALAAAREIEDEKRYARDALSLANKLPPELL--LSEAVAAVEIKDEKRYARATILSLAAKLP 913
QY 202 -----KSAIRSATAIOIGEOOKRFLPTLRNL 228
Db 914 LPEALAAAREIOFESSRAQALRSLADKLPPELLPEALAAAREIOBEYRAQ-----ALRSL 969
QY 229 LFSAPQODEALYALGKIKDQOSTYNNIKQL--OKPD--VDVTLAAQALALGKEEDAL 285
Db 970 ADKLPPELLPEVVAATREIEVEGEFRASIIIGSLAVQLPELLPEAVAAAREIEYDAFHSDAL 1029
QY 286 PVKKOALEERPRLAYALRH-----LPSELGIPALPIFLKTNSEAK 328
Db 1030 SNLAKKFEVLPPEALAAAREIIEHCRADSLSLADKLPPELL--LPEAVAAAREIODEYR 1088
QY 339 LNVALLLELGDGDPKLL--EYITERLYOPHYNETLALSFSGKRTLOMKNRANITVPDP 386
Db 1089 ---AVVILSLADKLPPELLPEAVAAARGIODEYPRSLAL-----RLAD-----KL 1131
QY 387 QERELLSTTGLEQOILTFLEPLKRAYLPCTYKYLASOKTQALATYATISFLSHSHEA 446
Db 1132 ELPEALAAAREIODD-----GHMRA 1152
QY 447 LDILFOAKLPGEPITIRAYADLAIYNLKDPKKRSLHDYAKKLIQETL 495
Db 1153 YNLSLDLADKLPPELLPKALA--AAAREIODEYRRANALSLADKLPPELL 1199

RESULT 11
S28294
hypothetical protein ZC84.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 12-Mar-1993 #sequence, revision 12-Mar-1993 #text, change 30-Sep-1993
R:Thomas, K.
submitted to the EMBL Data Library, December 1992
A:Reference number: S28285
C:Accession: S28294
A:Molecule type: DNA
A:Residues: 1-730 <THO>
A:Cross-references: EMBL:219157
C:Genetics:
A:Introns: 77/3; 120/3; 213/1; 228/2; 254/3; 332/3; 428/2; 499/3; 687/3; 727/3

```

Query Match 4.7%; Score 134.5; DB 2; Length 730;  
 Best Local Similarity 20.5%; Pred. No. 1.8;  
 Matches 139; Conservative 101; Mismatches 250; Indels 187; Gaps 27;

```

QY 1 MGFLHLLFLGL-----LCSLPISLVAKFPESVGHKILTYISTOSTOQALATYIEAL 51
Db 9 LGCELAVKDLRSQVLEAATCSFIVSKYGIETHSIGEDILV--PAMSQVAAVSTKIMAT 66
QY 52 DAYGDHDFEVLARKIGEDYLVKOSIHSSDPQTRKSTIIGAGLAGSEALDVLSSQAMETADPL 111
Db 67 SASLTTFEIV-----EYV-----QTR-----QVFTIILSFSTSKDKS 98
QY 112 QQLVLIVSAVSGHLGKTSDDLFLKALASPYPVIR-----LEAAVRLANKNTKVIDHLSF 166

```

```

Db 99 QRQLAALLLEIYISKWSDNIRKQIMROICELIKSAINDADSETRAAGRAFAKLDMHS- 157
QY 167 IHKLPEEIOCL-----SAAIFLRETESDAYIRDLAAKKSASIRASATA 210
Db 158 -----EAAIVLELDHOKMKIRGDDAASNAVSASEKSIPIRKISLBSAGSAHNHNTSA 212
QY 211 LOIGEQOKRFLPTLRNLITLSASPODEALYALGKIDQOS-----YYNIRKQLOKPD 264
Db 213 ISEKKGIKKLLPIYAD--TSLNPIEIKNIGCNLNLSDASNTWVLEIYSIFVRTHSSR 270
QY 265 VDVTLAAQALALGKEEDALPVKKOALEERPRLAYALRHLPESELGIPALPIFLKTN 324
Db 271 LSEWLRLALAKLPARKAAETLPNTKKQ-----IGHTLVNILECPN 310
QY 325 SEAKLNVALLLELGD-----TPK-----LLEYITERLYOPHYNETLALSFSGK-----R 370
Db 311 AHHQL--VTVCMLCDDPIHLMVPRARVVLLEYILSL--DEYTEFGASINNAKELKTAIR 365
QY 371 TLQNMKNRANITVPQPOQERELLSTTGLEOILTFLEPLKRAYLPCTYKYLASOK--- 427
Db 366 KMLTW-----ASDP--RLSILTPH--VERAKISMFCVNVADFSALISDIDSEQKNMI 414
QY 428 -----TOLATTAISFLSHSHEO--ALDILFOAKLPGEDI----- 461
Db 415 HOTLDRNGLENGISSNNIATNSGATASRSTSWTSFOKESTSGLPEFGARKGCTGVNIGS 474
QY 462 IRAYADLAIYNL-----TKDP-----EKKRSLHDYAKKLIQETLIFYDT 500
Db 475 LNISNNLALSRLEQSTSRLEKVMNLSTVLPPTLLEKIQVQD-----LLOKMMSSEA 530
QY 501 ENQRHPSMPYLR-----YQVTPESRTKLMDILELATSKSSED---IRLLQLMTEG 551
Db 531 DEQSAISSIYMMIDGGEGEWOCYAKLLNLFELLSRSRENNKMKCLRILGKMCYAO 590
QY 552 DAKNPPVLGLIKTIVE 568
Db 591 AAKLPDSTEMAVCKYLD 607

```

```

RESULT 12
C88550
protein ZC84.3 (Imported) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence, revision 10-May-2001 #text, change 10-May-2001
C:Accession: C88550
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C-
A:Accession: C88550
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-826 <STO>
A:Cross-references: GB:chr_III; PIDN:CA79568.1; PID:g3881445; GSPDB:GN00021; CESP:2C
C:Genetics:
A:Gene: ZC84.3
A:Map position: 3

```

Query Match 4.7%; Score 134.5; DB 2; Length 826;  
 Best Local Similarity 20.5%; Pred. No. 2.1;  
 Matches 139; Conservative 101; Mismatches 250; Indels 187; Gaps 27;

```

QY 1 MGFLHLLFLGL-----LCSLPISLVAKFPESVGHKILTYISTOSTOQALATYIEAL 51
Db 89 LGCELAVKDLRSQVLEAATCSFIVSKYGIETHSIGEDILV--PAMSQVAAVSTKIMAT 146
QY 52 DAYGDHDFEVLARKIGEDYLVKOSIHSSDPQTRKSTIIGAGLAGSEALDVLSSQAMETADPL 111
Db 147 SASLTTFEIV-----EYV-----QTR-----QVFTIILSFSTSKDKS 178

```

Query Match	4.7%	Score 134.5	DB 1	Length 1238
Best Local Similarity	20.1%	Pred. No. 3.8		
Matches 141	Conservative 101	Mismatches 227	Indels 233	Gaps 36

19 SLVAFPEPSVGKILIVISTSTQOALATYLELDAVGHDPFLAKIGEDVIKQSIHSD 78  
Db TLQAVYPOATHLYF-----PSSQALAAV-----AYGADVFIGALTSTTHSVSYFND 232  
OY 79 POT-----RSTTIGAGLASSBALV 101  
233 KRVVAPADQIVYGESFGVADNTLLRVYNAVLEAIPASERSLLYRML-GSSISLDP 291  
OY 102 SQAMETAD-----PLQQLLVLSVSGHLGKTSDDLFEKALASPYVIRLE----- 146  
Db 292 RPAYSAREQOMMANHPVAVLVAN-----LEAFETLFRTDEQGGISA 334  
OY 147 AAYRLANLK-----NTYVI--DHLSTFHKLPETIOCLSAIFLRLEEE-----SDAYIRP- 196  
Db 335 AVLQLOLRGTGDFDQIIGDVTEELIARSGEADMACALFNAAARESVLSFSRPVRNG 394  
OY 197 -LLAAKKSARSATLQIGEYQOKREFLPRLN-----LTSASPODO-----EAIY 242  
Db 395 LVIVTRQDPAPADA-----DHLDRGTIAMVRSAALPLIQOKYPOAKVYTADNPTEAML 450  
OY 243 ALGKLKDGQ-----SY-----NKKQLOKPDVDTLAA----- 272  
Db 451 ----VADQADAVVOTQISASYVNRYPAGKLRISALDLPPEATIALATARGOTELISL 506  
OY 273 -QALALALKEEDALPVIKQALEERRALYALR-----HLPSEIGIPALPIEL----- 320  
Db 507 NKALISISNDELASIVSRKSGSDGPRTWYAVARNREIYLLIGLLSAL-LFLSWIYVLYRR 565  
OY 321 ---KTKNSEAKNVALALLELGGD--TPKLLLEYTER-----VOPHYNETLALSEK--G 369  
Db 566 QIRQKRAERARNLNDLDEFRRVLDIGTPNPI-YVRKEEGMGLCNDAIYDTPGVTADEVIG 624  
OY 370 RLIQWKRNVNITVPODPOERERLLSTTRGLEQIILTFLEPRKEAVLPCTYLLASQKTQ 429  
Db 625 KVI---PEANVY--GDP-----ALAREMHFEFLT-----RMAAREPR 657  
OY 430 LATTATISFLSHSH--QEALDOLLFOAKLPGEPIIRAYADLAIVNLTQDKPEKRSIHD-- 485  
Db 658 FEDRDVTLGKTRHRYQWTPYIGDSLIGLKG--ITIGWID-----ITTEAEILLRELDHAK 710  
OY 486 -----YAKKLIQET-----LLEVDENORPHPSMFLRYQVTPESR 521  
Db 711 ESADANRAKKTFLATMSHEIRTPMNATIGMDELALLRPADQEPDQSIQVAYD--SARSL 769  
OY 522 TKLMDILETTLATSKSSED---IRLLIQMTGDAKNPVA 560  
Db 770 LELIGDIIDIAKIEAGKFDLAVFRTALRALPEGAIRVFDGLA 811

RESULT 14  
SI7946  
virulence sensor protein bvgS precursor - Bordetella parapertussis  
C:Species: Bordetella parapertussis  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: SI7946  
R:Ratko, B.; Scarlato, V.; Monack, D.M.; Falkow, S.; Rappuoli, R.  
MOL. Microbiol. 5, 2401-2491, 1991  
A:Title: Structural and genetic analysis of the bvg locus in Bordetella species.  
A:Reference number: SI7943; MUID:92167813  
A:Accession: SI7946  
A:Molecule type: DNA  
A:Residues: 1-1238 <ART>  
A:Cross-references: EMBL:X52948; NID:g39727; PIDN:CAA37124.1; PID:g39729  
A:Experimental source: Strain 9305  
A>Note: The authors translated the codon GAG for residue 134 as Gly  
C:Function:  
A:Description: Involved in signal transduction  
C:Superfamily: eysg protein; response regulator homology  
C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferas  
F:1-30/Domain: signal sequence #status predicted <Sig>  
F:31-1238/Product: virulence sensor protein bvgS #status predicted <Mat>  
F:698-945/Domain: sensor histidine kinase homology <Shk>



F.795-1091/Domain: response regulator homology <RRH>  
F.729/Active site: His (phosphohistidine intermediate) #status predicted  
F.1023/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match	4.7%;	Score 134.5;	DB 1,	Length 1238;
Best Local Similarity	20.18;	Pred. No. 3.8;		
Matches 141;	Conservative 101;	Mismatches 227;	Indels 233;	Gaps 36;
QY 19	SLVAKFPESVGHKLLIYSTOSTQALATYTEALAYAGDHDFVLRKIGEDYKQSIHSSD	78		
Db 183	TLQAYAPQATLHYF-----PSSQALAAV-----AYGADVIFIDALLTHSLVQSQSFND	232		
QY 79	PQT-----	101		
Db 233	VRVVAPOIVTGESFGVRAADNTRLRLRVNAVLEAIPASERSLRSLYRMGL-GSSISLDF	291		
QY 102	SQAMETAD-----PLQQLLVLSAVSGHGLKTSDDLLFKLASPPYVIRLE-----	146		
Db 292	RPATSAHQOMAHNPVAKVAVLN-----LEAPFLFRTDEQFGGISA	334		
QY 147	AAVYRLANLK--NKKVI--DHLHSFIKRLPEEIQLSAIFRLPETEE---SDAYTRD-	196		
Db 335	AVIQLQLRTGLDPIQIIGVDIVLEELIKLRSGEDMAGALEFVNAAKRSYLSFSKPYRNG	394		
QY 197	-LLAAKSAIRSATALQIGEYQOKRFLPTLRN-----LTSASPOD-----EALY	242		
Db 395	MVIYTRDDPAAPADA---DHLDRITAMVNSAIPILQOKRYQAKAVYADNPTTEAMLL	450		
QY 243	ALGKLKGG-----SYX-----NKKQLOKRPDVVITLAAA-----	272		
Db 451	---VAQGQADAVAVQTOISASYVNRYPAGKRLRTASALDLPAPALATARGQETLSIL	506		
QY 273	-QALIALGKEEDALPVTKQALEEPRALAYLR---HLPSIEGIPALPLFL-----	320		
Db 507	NKALYISNDELSIVSRWKSODGPWTWYAYRNEIYLLIGLISL-LFLSMYIVLR	565		
QY 321	---KTNKSEAKLNVALLALEGCD-TPKLELEYTER-----LYVQPHYNETLASFSK-G	369		
Db 566	QIQRKRAERERALNDQLEFMRLIDGTPNPI-YVADKEGRMLCNDVADLTGFGVADAVLG	624		
QY 370	RTLOMKNRNVINIVQDOQERELLSTRGLEQLTLFLRLPKFAYLPCTIKYLLASQTKQ	429		
Db 625	KTI--DEANV--GDP-----ALARENHFEFLT-----RNAAREPR	657		
QY 430	LATTAFISFLSHSH--OEALDLEQAALGPEPTIRAYADLAIYNTLKDEPKKSLHD--	485		
Db 658	FEDRDVTLHGKTRIVYQWYVPYDGSGLGKG--ITGSMID-----ITERELLRELHD	710		
QY 486	-----YAKKLIQET-----LLFVDTENQRPHPSPYLRYQVTPESR	521		
Db 711	ESADANRRAKTTFLATSHETIRPMNAIIGMLELALRLPADQEPDRDSIQVAD-SARSL	769		
QY 522	TKMLDLLETFLATSKSED---IRLLQLMTBEGDAKNFPVLA	560		
Db 770	LELIGLIDILAKIEAGKFDLAPVTALRPEGAIRLFDDGA	811		

RESULT 15  
T37768  
probable vacuolar biogenesis protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_rev10n 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T37768  
R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
Submitted to the EMBL Data Library, June 1997  
A:Accession: T37768  
A:Reference number: Z21738  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-860 <MUR>  
A:Cross-references: EMBL:297185; PIDN:CA80996.3; GSPDB:GN00066; SPDB:SPAC16A10.03a  
A:Experimental source: strain 972h; cosmid c16A10

C;Genetics:  
A;Gene: SPDB:SPAC16A10.03cc  
A;Map position: 1  
A;Introns: 40/3; 100/3

Query Match	Similarity	4.7%	Score 134	DB 2	Length 860
Best Local	Similarity	18.4%	Pred. No. 2.4		
Matches 118	Conservative	85	Mismatches 234	Indels 206	Gaps 19
OY	12	LLCSLPISLVAKFPSPVSGHKILYISTOSTOYALATY-----LEAL-DAVGADHDFVLRLKI	65		
Db	228	LECS--IYOGKYMICAGGSFLSYVTTPTDMLQNTYCVDGFPRELFSSGFLVYVYTRKN	285		
OY	66	GEDYLKQSIHSDPOTKRSTIIGAGLACSSSEALDVLSOAMEFADPVLQOVLIVSAVSGHIG	125		
Db	286	GEN-----GLENNSSIREIKALDVERRYVLYESLSEQSDNIFE	324		
OY	126	KTSDDLLEKALASPYVIRLEAAYIRLANLKRVVDIHSHSFHKRPREEIOCSAAIFRL	185		
Db	325	NSFDCIFFSSFTVPQQLIRLPDFFVLCKMKGG-----	357		
OY	186	EYESDAY-IRDLAAKSATSIRSATADIGE-YQQRKELPTLRNLITSASPODEALILYA	243		
Db	358	--EHKDKARKIANYLSPEDETIRECALAAAGECROOLNF-----ODATYUY	400		
OY	244	LGLTKDGGSYNNIKQLOKPDVDVTLAAAOALIALG-----	279		
Db	401	IEAIFPSSAETIKRYLEKKLIKELTSLVEALSAAGFAPSHETSLIYLIVLYIKRLDKLT	460		
OY	280	-----KEEDALPVTKKQALEERPRALYALRHLPs-----EIG-----IPIA-LP	317		
Db	461	EYVSGCPTEISIPILIRKCYKCDOMELLCTIKRLPVCMEVYQOEKDDVEKAENHLOVCMIP	520		
OY	318	IFLKRRNS-----EAKINVALALLEGGDPFKL----	345		
Db	521	ELKRTSNSEFGVILWNSDPMREPKAKIRINIEIINSGCKREKSLNLIKTYLGLFSSONVOIQ	580		
OY	346	LEYITERVOPHYNETTALSEFSKRGTTLONMKRVNIIVPODERERRLSTTGRGEEOILT	405		
Db	581	LIFDELLEKSKSENVLFYTRKLYALMOKELQSNPNEDLADQIITHDSGL-----	634		
OY	406	FLFLRPKRAYLPCLYKLLASQKTQATATAISLSTSHQEOALDILFQAAKLPGERTIYAY	465		
Db	635	-----LDYESSITLCQAVSWKOVTDL-----LYSHLSLKEGOD-----	667		
OY	466	ADLAAYNTKTPPEKRSJLHDYAKKLIQETLLFVDENORPHRSMPLYRQYVPESTRKLM	525		
Db	668	-DSLQOQIISDPEYKTL-----SETYSDEALHVLKFFVRERSITMKY	710		
OY	526	LDIETLATSKSSF--DIRLLIQMTGEGDAKNPVLAGGLIK	565		
Db	711	EDLYKRIEACFMQFRIPICQVNLIVADGTLNFCFLKPLLIK	753		

Search completed: August 20, 2002, 04:12:03  
Job time: 3068 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 20, 2002, 01:14:10 ; Search time 2366.15 Seconds  
(Without alignments)  
16865.755 Million cell updates/sec

Title: US-09-662-812-1

Perfect score: 1907  
Sequence: 1 gtgctgtgatttgaaaaaag.....ctcatcaagagcgtgtgaa 1907

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 segs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_ov.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	1907	100.0	1907	6	AX100528	AX100528 Sequence
2	1907	100.0	8268	1	AE002234	AE002234 Chlamydia
3	1907	100.0	16448	1	AE001587	AE001587 Chlamydia
4	1707	89.5	1707	6	AX349527	AX349527 Sequence
5	84	4.4	2088	6	AX349569	AX349569 Sequence
6	24	1.3	36	6	AX100531	AX100531 Sequence
7	24	1.3	44	6	AX100530	AX100530 Sequence
8	22	1.2	11627	1	AE001308	AE001308 Chlamydia
9	22	1.2	47445	2	AC017438	AC017438 Drosophila
10	22	1.2	183722	3	AC009341	AC009341 Drosophila
11	22	1.2	261000	3	AE003654	AE003654 Drosophila
12	21	1.1	4343	6	AX035433	AX035433 Sequence
13	21	1.1	177014	2	AL670958	AL670958 Mus muscu
14	21	1.1	206591	2	AL611984	AL611984 Mus muscu
15	21	1.1	225609	2	AL645546	AL645546 Mus muscu
16	20	1.0	558	10	MUSRABAP	MUSRABAP Mus muscu
17	20	1.0	2680	5	AMWNTIR	AMWNTIR Mus muscu
18	20	1.0	3330	8	AF024648	AF024648 Arabidops
19	20	1.0	3620	10	BC022119	BC022119 Mus muscu
20	20	1.0	4907	10	MMU245569	MMU245569 Mus muscu
21	20	1.0	17854	1	AF307053	AF307053 Thermococ
22	20	1.0	18001	1	AF307052	AF307052 Pyrococu
23	20	1.0	23619	2	AC102470	AC102470 Mus muscu
24	20	1.0	35857	2	AC026146	AC026146 Homo sapi
25	20	1.0	52347	2	AC068214	AC068214 Homo sapi
26	20	1.0	68996	2	AC102790	AC102790 Homo sapi
27	20	1.0	86212	2	AC106082	AC106082 Rattus no
28	20	1.0	104534	9	AL356317	AL356317 Human DNA
29	20	1.0	104944	9	AL591509	AL591509 Human DNA
30	20	1.0	105168	9	AC091317	AC091317 Homo sapi
31	20	1.0	114567	9	AC010473	AC010473 Homo sapi
32	20	1.0	124146	9	AC068706	AC068706 Homo sapi
33	20	1.0	131691	9	AC096765	AC096765 Homo sapi
34	20	1.0	139616	2	AL445261	AL445261 Homo sapi
35	20	1.0	146935	9	AC096647	AC096647 Homo sapi
36	20	1.0	148845	9	HS114A1	AL035684 Human DNA
37	20	1.0	149671	9	HS376D21	Z98946 Human DNA
38	20	1.0	151976	9	AC024588	AC024588 Homo sapi
39	20	1.0	156785	2	AC018541	AC018541 Homo sapi
40	20	1.0	164394	2	AC022873	AC022873 Homo sapi
41	20	1.0	164819	2	AC092335	AC092335 Homo sapi
42	20	1.0	166692	2	AC023475	AC023475 Homo sapi
43	20	1.0	167332	2	AC080168	AC080168 Mus muscu
44	20	1.0	170736	2	AC013658	AC013658 Homo sapi
45	20	1.0	170933	2	AC087619	AC087619 Homo sapi

#### ALIGNMENTS

RESULT 1  
AX100528  
LOCUS AX100528 1907 bp DNA linear PAT 10-APR-2001  
DEFINITION Sequence 1 from Patent WO0121804.  
ACCESSION AX100528  
VERSION AX100528.1 GI:13619532  
KEYWORDS

#### SOURCE

Chlamydia pneumoniae.

#### REFERENCE

1 (bases 1 to 1907)  
Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.

#### TITLE

I Chlamydia /I antigens and corresponding dna fragments and uses thereof

#### JOURNAL

Patent: WO 0121804-A 1 29-MAR-2001;

#### FEATURES

source

1..1907  
Location/Qualifiers

/organism="Chlamydia pneumoniae"

/db\_xref="taxon:83558"

/note="unnamed protein product"

/codon\_start=1

```
/transl_table=1
/protein_id="CAC36452.1"
/db_xref="GI:13619533"
/translation="MGLEFHLTFLGLLCSLPIISLVAKPESVGHKLIYISQSTQOAL
ATYLEADAYGDHDFVPLRKIGEDYLKOSIHSDPQRKSTIIAGAGLSSSEALDVL
QAMETADPLQDLVLVSAVSHGLKTSIDLEKALASPVPVRLSAAYLAKMKNKI
DHLSFTHKLPDEIQLSAAIFLELRESEAYIRDLAAKRSIRATLOIEYOO
KRRLPTLRLLTSASPODEAILYALSKLKGQSYNINRKLQKPDVDTLAAAOALI
ALKEEDALPVTKQALBERRALYALHLPSEIGIPALPFIKTKNSEKLVVALA
LLEIGDTPKLEYITERLVQPHNLETSFSGRLLQNMKRVNIIVPODERERL
LSTRGEBEOLTFELFRLPKFAYLPCYIKLASOKTOLATISFLSTSHSEALDL
FOAKLPGEPILIRAYADLAIYNLTDEKSKLSDYAKKLTQETLFLVDTEPNRPHS
MPYRYQVTPESRTRKMLMDILETLATSSSDIRLLIQLMEDGAKNPVLAGLLIKI
VE"
BASE COUNT      542 a      456 c      380 g      529 t
ORIGIN
Query Match      100.0%; Score 1907; DB 6; Length 1907;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtggcgtatttggaaaaagtcacatgagatggttataatgltcaaggtctccatcc 60
Db 1 GTGCGTGTATTTGAAAAAGTCATGATGTTTATATGTTCAAGTCTCCCTATCC 60
QY 61 aaacttgaactactgcttagagaggttgaacatcgatctatggaactatccatcaac 120
Db 61 AAACCTTGAATACTTGTGTAGAGGATGTAACATGATCTATGAGCATTCATCTAAC 120
QY 121 tctcttggaacttattggttagcttcccatctctctctgttctgtaattcccgagtc 180
Db 121 TCTCTTGGAGCTTTATGTGTAGCTTCCATTTCTTGTGTTAAATTCCTCGAGTC 180
QY 181 tgaagtcataagaccccttataatgaagacaaatcaacagcggccttagacaata 240
Db 181 TGTAGGTATTAAGATCTTTATATATAGCAATCTAACAGCAGCGCTTAGCAACATA 240
QY 241 tctggaagctctagatgctcagatgacatgactctctctgtttaaagaataatcgaga 300
Db 241 TCTTGGAAGCTCTAGATGCTTACGGGTGATCATGATCTTCTGTTTAAAGAAAAATGAGAGA 300
QY 301 agactatctcaagcaaaagcaccctctcagatccgcaaaatagaanaagcaccatcat 360
Db 301 AGACTATCTCAAGCAAAAGCATCCACTCTCAGATCCGCAAACTAGAAAAAGACCATCAT 360
QY 361 tggagcaagcctggcggagatcttcagaagccttggacgtgtcttcccaagctatggaac 420
Db 361 TGGAGCAGGCTTGGCGGAGATCTTCAGAGCCTTGGAGCGTGTCTCCCAAGCTATGGAAC 420
QY 421 tgcagaccccttgcagcagcagctacgtgttatacggcagctcaggaacatcttgggaaac 480
Db 421 TGCAGACCCCTGCGAGCAGCTACGTGTTTATCGGCAGCTTCAGAGCATCTTGGGAAAAAC 480
QY 481 ttctgaagcactactggtttaagccttagacatcccatctcgtatcagcgttagaagc 540
Db 481 TTCTGAGCAGCTTACTGTTTAAAGCTTTTATGCACTTCCTATCTGCACTCCCTTAGAAGC 540
QY 541 cgcctatagactgtctaaattgaagaacactaaagltcatltgacatctacatcttcttcat 600
Db 541 CGCCTATAGACTTGTCTATTTGAAGAACACTAAAGTCAATGATCATCTACACTTCTTTCAT 600
QY 601 tcaataagctccgagaagaatccaatgctcatctgcggaacatctcctaagccttggagac 660
Db 601 TCATAAAGCTTCCCGAAGAAATCCAAATGCTATCTGCGGCAATATTTCCATCTTGAGAAC 660
QY 661 tgaagaatctgaigtctatatctcgagatctcttagctgccaagaagaagcgagatctcgag 720
Db 661 TGAAGAATCTGATCTTATATCTGGGATCTCTTAGCTGCGCAAGAAAGCGGATTCGGAG 720
QY 721 tgcacagccttgcagatcgcgaataccaacaaacgcttctctcgacacttagaa 780
Db 721 TGCCACAGCCTTGCAGATCGGAGATACCAACAAACGCTTCTTCCGACACTTAGGAA 780
```

```
QY 781 ttctgtaacgagtcgctctcctcaagatcaagaactcttcttctgctttagggagct 840
Db 781 TTTGCTAACGAGTGGCTGCTCCCAAGATCAACAAACATTTCTTATGCTTTAGCGAAGCT 840
QY 841 taagatgtgcaagcctctacatatataaagaatctggaagcctgagtgagtg 900
Db 841 TAAGATGTGTCAGACTCTACTCAATATATAAAGCAATTTGCAAGAGCTGATGTGATGT 900
QY 901 cactttagcagcagctcgaagcttcaatgtcttgggaaagaagcagcctctccgct 960
Db 901 CACTTTAGCAGCAGCTCAAGCTTTAATGTCTTTGGGAAAGAGAGAGAGCGCTCTCCGCT 960
QY 961 gataaaaaagcagcacttggagagcgccctcgagcctctgtatgctccttaagcactacc 1020
Db 961 GATFAAAAAAGCAAGCACTTGAAGAGCGCGCTCGACCCCTGTATGCTTACGCACTAC 1020
QY 1021 cctcgaagaagagatcccgatcgccctccgcatctccataaagaagaagaagc 1080
Db 1021 CTCTAGATAGGAGATTCGATTCGATTCGCTCGCATATTTCTTAAACATTAAGAACGGAAGC 1080
QY 1081 caagttgaatgtagctttagctctcttagagtgtaggggtgtagcacccctaaactactgga 1140
Db 1081 CAAGTTGATGTAGCTTTAGCTCTCTTAGAGTTAGGAGGTGACACCCCTAAACATCTGGA 1140
QY 1141 atacattaccgaagagctgtcccaacacatatataatgaagacttaagccttgaagttctc 1200
Db 1141 ATACATTTACCGAAAGGCTTGTCCAAACACATTTATATAGACTTATGCTTATGAGTTTCTC 1200
QY 1201 taaaggcgctactttagaanaatttgaagcgggttgaacatgaagtcctcctaagaatcccca 1260
Db 1201 TAAAGGCGCTACTTTACAAATTTGAAAGCGGCTGAAACATCATATGCTTCAAGATCCCA 1260
QY 1261 ggaagggaagaagtgctctccacacacacagaggtcttgaagaagcagatcttact 1320
Db 1261 GGAGAGGGAAGAAGTGTCTCTCCACAAACCGAGGCTTGAAGAGCAGATCTTACCTTTCT 1320
QY 1321 ctccgcctacctaagaagacttaccctccctgtattataagcttcttggcgagtcagaa 1380
Db 1321 CTTCCGCTACTTAAGAAAGGTTTACCTCCCTGATTTATATAGCTTTGGGAGAGCAAGAA 1380
QY 1381 aactcagcttgcacactagctgattctcttctttagagacacccctacaatcaggaagcctt 1440
Db 1381 AACTCAGCTTGCCACTACTGCGGATTTCTTTTATAGTCAACCTTCACATCAGAGAGCCTT 1440
QY 1441 agatctacttctcgaagcttgcgaagcttcccttggagaaccttcatcgcgcgctatgcaga 1500
Db 1441 AGATCTACTTTTCCAAAGCTGGGAAGCTTCTGGAGAACCTTATCATCCGCGCTATGACAGA 1500
QY 1501 tcttgctattataatctcaccacaaagatccttgaaaaaacgcttctcccatgattatgc 1560
Db 1501 TCTTGCTATTTATATATCTCACCAAAAGATCTGAAAAAAAGGTTCTTCCATGATATATGC 1560
QY 1561 aaaaaagctaatcagaagaaccttgttattgttgaacacggaacaaacaaagaccccatcc 1620
Db 1561 AAAAAAGCTAATTTACAGGAACCTTGTATTGTGACACGGAACCAACAAAGCCCATCTCC 1620
QY 1621 cagatgcccctatctaglttcaaggtcaccccaagaagcgttagcgaagctcatglttga 1680
Db 1621 CAGCATGCCATCTAGCTTATCAGGTTCACGCCCAAGAAAGCGGTAGAGAGCTCATGTGGA 1680
QY 1681 tatcttagagacactagcaccctcgaaagcttccggaagatatccgttattatgatacaact 1740
Db 1681 TATCTTAGAGACACTAGCCACCTGGAAGTCTTCGGAAGATATCCGTTTATGTATGATCAACT 1740
QY 1741 gataggaagaagatgcaaaaaatctcccgagctccttgcagagcttaccatataaataatgt 1800
Db 1741 GATAGCAGAAAGAGATGCAAAAAATTTCTCCAGTCTTGCAGGCTTACTCATATAAATTTGT 1800
QY 1801 ggaataaccccaactacgtcttcttgaagaagcttgccttcttcttcttagcttcccttgc 1860
Db 1801 GGAATTAACCCCAACTACGCTTATGAAAGCTTGCCTTATTTCTTACGCTTCTTTGTTC 1860
```

Qy 1861 ttatgggttcctcaagctatgcttctgcatcatcaagagctctgaa 1907  
|||||  
Db 1861 TTTATGGTTCCTCAGCGATGCTTTTGACTATCATCAAGAGCTGTGAA 1907  
|||||

RESULT 2  
AE002234 8268 bp DNA linear BCT 30-MAY-2000  
LOCUS Chlamydomophila pneumoniae AR39, section 62 of 94 of the complete  
DEFINITION genome.  
ACCESSION AE002234 AE002161  
VERSION AE002234.2 GI:8163492  
KEYWORDS  
SOURCE Chlamydomophila pneumoniae AR39.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
REFERENCE 1 (bases 1 to 8268)  
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.,  
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39  
Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
PUBMED 10684935  
JOURNAL MEDLINE  
2 (bases 1 to 8268)  
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,  
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,  
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.,  
Direct Submission  
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
Medical Center Dr. Rockville, MD 20850, USA  
On Jun 1, 2000 this sequence version replaced gi:7189667.  
LOCATION/Qualifiers  
1..8268  
/organism="Chlamydomophila pneumoniae AR39"  
/strain="AR39"  
/db\_xref="taxon:115711"  
/note="synonym: Chlamydia pneumoniae AR39"  
113..1036  
/gene="CP0751"  
113..1036  
/note="similar to GB:U00096 SP:Q47012 PID:1381661  
PID:178603; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="AtsA/FlaC family protein"  
/protein\_id="AF38556.1"  
/db\_xref="GI:7189668"  
/transl\_table="MSSRELLIIIGSSQOQPTRTNRQAGVLFNMGEGELFDPEQGTOR  
QTFANIAPTVNRIVSFHSHGSDHCLGSLMRLNLDVSHPIHCYVYASGKVTOR  
LNGTIVHEITIOVEHPISEEGIVDEGSEFTEAORLQHOVDLIGRITPEPTIKLP  
KLESGRIGLIIODLIRPOEISIGSSTYVLSVSVYRKGDSTAIITADLLPCQAALDD  
AKNSCMCESTYLEDQHRHLAESHFMHTAKQATLAKRAATQKLIIILHFSARYLNDL  
FYKEASAVPEPNVSAOERYSPPKPNPLNK"  
1121..2059  
/gene="CP0752"  
1121..2059  
/note="CP0752"  
/note="Integrase/recombinase, phage integrase family"  
identified by match to Pfam protein family HMM PF00589"  
/codon\_start=1  
/transl\_table=11  
/product="Integrase/recombinase, phage integrase family"  
/protein\_id="AF38557.1"  
/db\_xref="GI:8163493"  
/transl\_table="MTASITYSFLDYLKWKSASPHTLRNYCIDLNLGKIFLEERGNIA  
PSSPIOLATEKRVSELPFSLFTEKHEVRMYIAKLINCGAKAKRTIKRCLSSIKSFAYHC

gene  
CDS  
2114..3706  
/gene="CP0753"  
2114..3706  
/gene="CP0753"  
/note="ABC transporter, ATP-binding protein; identified by  
match to Pfam protein family HMM PF00005"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter, ATP-binding protein"  
/protein\_id="AF38558.1"  
/db\_xref="GI:8163494"  
/transl\_table="MSTVLDKICKSLGCTRIPLDDVSVENPCGCTGTPENGAKSTL  
LKIKMGMEPTRGSTISLPKRVGILRQNIIDSFHTVLDVCYIMKNTIRMEALQRRNLY  
LQFTDAIGMEIEIEIIGENGVRADSAEELIGIGIPNMPDKKAMIPIDLOF  
RYLCOALGHPALLIDEPNTNLDLISINMLKFLKDEGVIVVSHDRHPLNTIT  
HIAIDYDITIIYPGNVDVMEKRTASROEKADIKSEKKSISQLEPYAKPGAGSRA  
SOVQSLREIKKIQPOEIKSNRQRYIRPISDKSGKVIISLAITQDYGHOVYH  
PSLEIYQDKLIGTGNNGIGKTLTKLAGVAPBSGSIKIGHOAICSTFPONHSDV  
LADCGEETLEFWLNRKKTGINDEIRSVLGKMLEGGDDAFKOIALSGGETARLLAG  
MLENHNVLILDEANNHLDLESVSALSMIAINDYKTAIFVSHDRGLIDCATKLLIFD  
KDKITFPFGTMVDTYTAGHKOLL"  
3735..4325  
/gene="CP0754"  
3735..4325  
/gene="CP0754"  
/note="similar to GB:L08793 SP:Q02169 GB:M96343 PID:142853  
PID:143161; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="mat protein"  
/protein\_id="AF38557.1"  
/db\_xref="GI:7189669"  
/transl\_table="MSLPYLGSSPRRKFIIEKFPVPTVIPSNDKSVSGDPI  
AYNGLAOKAVAVSELSPCDIIITGDTIVYDGRIFEPQDKADAIOMLKTANO  
THDVYSTAVYHKGKLTGSETSQISITMPDRISYIDYVGTLLNCGAIVDCRGL  
IKKHVGCYVNVGLPIQTLKYLEELNDLMDYST"  
4307..6013  
/gene="CP0755"  
4307..6013  
/gene="CP0755"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AF38558.1"  
/db\_xref="GI:7189670"  
/transl\_table="MGLFHLTLFGLLCSPLISLVAKFPESVGHKILYISTQOAL  
ATYLEALDAVGDHDFVLRKIGEDYILKQSIHSSDPQTRKSTIIIGAGLAGSEALDVS  
QAMETADPLDQLLVLSAVSGHLEKTDSDLFKALASPVYIRLEAVYRLANLNTVYI  
DHLHSFIRLKPBEIIOCLSAIFRLTEESDAIRDLAKSAISAYATLQAGEYQ  
KRLPLPLRNLTASPODPAIIYALGKIKDGQSYNNIKKOLQPPDVATLAAQOII  
ALGKEDPAIVKQALREBRPAIVLRLHPSGIGRIAPILKIKNSKAPKANVLA  
LTELCDTPKLEIYTERLNQPIHNETLALSFKGRTLOMKAVNIIIVDDPDERRL  
LSTTKLEDEITLFLRLPREAVLPCTIKLASQKQTLATTAISFISHSHOARDLL  
FOAKLPGBEIIIRAVADLAIVNLTQDPEKRSADYAKKLIQETLLEFVDENRPHS  
MPYLRQVTPESRTRKIMLDILETLATSKSSSEDYRLLIQMLTEGDARNFVLAGLLIKI  
VE"  
6003..8117  
/gene="CP0756"  
6003..8117  
/gene="CP0756"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AF38559.1"  
/db\_xref="GI:7189671"

/translation="TMSNPRLMKRCFLPLASFLVINGSSADALTTHOEAUKKNSYLS  
HFKVSGIVTDEDVLNHNLRIOANKVYVENGOSIKIVANGNVNRAKTLVC  
DYLEYEDTDSCLLTNGRFAMYPWFLOGSMITLPELIVIKGITISEGPKDCLJS  
GDYLEYSSDSLISGKTLTLYRCLIPILFPEFSIMPEIIPKPIINFGSGFLGSLYL  
GMSYSPISRKHFSTFLDSEFKHVGMEFLHCSQOVENPMNMSYAHRLAIDM  
AEADRYRLHGDCEFTFHDHVNSEGEHLSDPEWADLFEPMFLKNTMGPRVCTNM  
DNEEGYLTSSVKNSFONANOELPYLLROYPSIINTGYLENIYECGYLANAFSD  
HYVENESSRLIARAPKLHKTYPLPICTISSTIGSSLIYSDVPEISRSQSALKD  
LDYFELHKSIIQORHIIPEPVITETTERPLAKNEDYIESIUDAFHSLNLLKAGIDT  
SVLSKINPREPRIHAKLMTHTILSTNSKPEPKYACELSLPFKNKNTVSLDAEMIK  
KHCHDMNIRHEWIGDNVAMTLESLSRHSKSLKCDRENILIDVSRPIDOLDSPLS  
DHRMLILGKLFVRPHPCMNRYLSLRGWHRODPNVEYQIILSTKLEFHMQLYGYVE  
RREADSEFFFLKDKPKKPE"

BASE COUNT 2399 a 1827 c 1667 g 2375 t  
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 8268;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtggcttgatttgaagaaagtcacatgagatggttataatgltcaagatccctacc 60  
DB 4207 GTGGCTTGATTGAAAAAGGTCATGATGTTTATATAGTCAAGGTCCTCATCC 4266  
QY 61 aaacattgaacttctgtagagagattgaacatcgatctatggactattccatctaac 120  
DB 4267 AAACATTGAATACCTTGCTAGAGAGTGGAACATCGATCTATGGACATTTCCATCTAAC 4326  
QY 121 tctcttggaactttatgtgttagtcttccatcttcttcttgcttaattccctgaatc 180  
DB 4327 TCTCTTGAGACTTTATGTGTAGTCTTCCATTTCTTTGCTTAATTCCTGAGCTC 4386  
QY 181 tgtatgataagaatcccttatalataaglacgaaatctacacagagccttagacaata 240  
DB 4387 TGTAGGTCATATAGATCCCTTTATATAGTACCAATCTACACAGAGCCTTAGACAATA 4446  
QY 241 tcttgaagacttaagatgctcctaggtgatcatgaacttcttctttaaagaataatcgaga 300  
DB 4447 TCTGGAAGCTTAAATGCTTACGCTGATGATCATGACTTCTTCTTTAAAGAAAAATCGGAGA 4506  
QY 301 agactatctcaagcaaaagcatccactccctcagatcccgcaaaatagaaagaacacatcat 360  
DB 4507 AGACTATCTCAAGCAAAAGCATCCACTCCTCGATCCGCAAACTGAAAAAACCACCATCAT 4566  
QY 361 tgaagcagagccttgcgagatcttcaagaagccttgaacgtgtctctcccaagctatgaaac 420  
DB 4567 TGGAGCAGAGCCTGCGGAGATCTTCAGAAAGCCTTGAGAGTGTCTTCCCAAGCTATGGAAC 4626  
QY 421 tgcagagccccctgagagagcactggtttatgcgacgtctcagaagaatctcttgggaaac 480  
DB 4627 TGCAGACCCCTGCAGCAGCTACTGTTTTATTCGCAAGTCTCAGGACATCTTGGGAAAC 4686  
QY 481 tctctgaagacttactgtttaaagctttagcatctccctatccctgtcatcgcgttagaagc 540  
DB 4687 TTTGAGAGACTTACTGTTTAAAGCTTTAGCATCTCCCATCTCTCATCCGCTTAGAAGC 4746  
QY 541 cgcctatagacttgctcaatttgaagaacaataaagtcattgatactcatcttctcat 600  
DB 4747 CGCTTATAGACTTGTCTAATTTGAAGAACACTAAAGTCAATGATCATCTACTCTTCAT 4806  
QY 601 tcataaagcttcggaagaataaccatgctatctgcgagcaatatctcttaagcttggagac 660  
DB 4807 TCATAAGCTTCCGGAAGAAATCCAAATGCTATTCGCGCAATATTTCTTACCTTGGAGAC 4866  
QY 661 tgaagaacttgatgctataatctcgagatctcttagctgcgcaagaagaagcagatctggag 720  
DB 4867 TGAAGATTCGATGCTTATATTCGGGATCTCTTAGCTGCCAAGAAAAGCCGATCTCCGAG 4926  
QY 721 tgcacagagcttgcagatcgagaataacacaanaaagcttctcttcgacacttagaa 780  
DB 4927 TGCCACAGCTTTGCGAGATCGGAATACCAAAAAGCGTTCTTCCGACACTTAGGAA 4986

QY 781 ttgttaacagagtggtgtctctcccaagaatcaagaactatcttcttgcttagggagact 840  
DB 4987 TTTGTAAAGAGTGGCTCTCCCAAGATCAAAACATATTTTATGTGTTTAGGAAGCT 5046  
QY 841 taagatgttcaagactactacatatataaaagaacttgaagagcctgagatgtagatg 900  
DB 5047 TAAGATGTCTCAGACTACTACTAATATATAAAAAGCAATTTGCGAAAGCCTGATGTGATGT 5106  
QY 901 caatttgaagcagcttcaagctttaaatttgcttgggaaagaagaagcagctcttccgt 960  
DB 5107 CACTTACAGAGCTCAAGCTTTAATGTCTTTGGGAAAGAAGAGAGAGCGCTCTCCCGT 5166  
QY 961 gataaaaaagcagcacttggagagcggtcctgagcctgtatgcttaagcacttac 1020  
DB 5167 GATAAAAAAGCAAGCATTGAGAGAGCGGCTCGACCCCTGTATGCTTACCGCATCTAC 5226  
QY 1021 ctctgagataggaattccgatttgccttcgcatatcttccataaataagaagaagagc 1080  
DB 5227 CTCTGAGATAGGATTCGCGATTGCTGCTGCGATATTCTTAANAATTAAGAACAGCGAAGC 5286  
QY 1081 caagltgaatgttagctttagctctctttagaglttaggtgtgacacccctaaactacttga 1140  
DB 5287 CAAGTTGATGATGACTTTAGCTCTCTTAGAGTTAGGATGTGACACCCCTAAACTACTGGA 5346  
QY 1141 atacattcggaaagccttgcacacacatataa tgaagacttagccttgagttctc 1200  
DB 5347 ATACATTACCGAAAGGCTTGTCCAAACCAATTATATGAGACTTAGCCTTGAGCTTCTC 5406  
QY 1201 taaaggggtacttcttaagaatttgaagcgggtgaacatcattagccttaagaatcccca 1260  
DB 5407 TAAGGGGCTACTTTACAAAATTTGAAAGCGGATGAACATCATAGTCCCTCAAGATCCCA 5466  
QY 1261 ggagagggaaaggttgcctccacacacagaggtcttgaagaagcagatccttact 1320  
DB 5467 GGAGAGGAAAGGTGTGCTCTCCACAAACCGAGCTTGAAGAGCAGATCTTAGCTTTCT 5526  
QY 1321 ctccgcctactaaagaagcttaccctccctgtatattaaagcttttggcgagtcaga 1380  
DB 5527 CTTCGCTTACTTAAGAAAGCTTACCTCCCTGTATTTATAGCTTTTGGGAGTCAAGAA 5586  
QY 1381 aactcagcttgcactacttagatcttcttctttaaagtaacaccccaacaaagagcctt 1440  
DB 5587 AACTCAGCTTCCACTACTGCGATTTCTTTTAAAGTACACCTCACATCAGGAAGCCTT 5646  
QY 1441 agacttacttccaaagcttgcgaagcttccctgcggaaccta tcalcgcgcgactatgcaga 1500  
DB 5647 AGACTACTTTTCCAAAGCTGGAAGCTCTCTGGAAGACCTATCATCCCGCTATGACAGA 5706  
QY 1501 tctgtcatattataatctcaccacaaagatccctggaaaaaaagcttctccatgatatgc 1560  
DB 5707 TCTTGCTATTTTATATCTCACCAAAAGATCTGAAAAAAGCTTCTCCATGATTATGC 5766  
QY 1561 aaaaaagctaatcagaagaacctgttatttgcgacggaagaaaccaaagaccatcc 1620  
DB 5767 AAAAAAGCTAATTCAGGAACCTTGTATTGTGTGACACGGAAAAACCAAAACCCCATCC 5826  
QY 1621 cagcatgcctatctacgltatcagltacagltaccccaagaaagccgtacgaagctcatgttga 1680  
DB 5827 CAGCATGCCCTATCTAGCTTATCAGGTACCCCGAAGAAAGCGTACGAAGCTCATGTTGGA 5886  
QY 1681 tatcttagagacacttagccacactcgaagcttccgaagatatccgttatattgataaact 1740  
DB 5887 TATTCTAGAGACACTACCCACTCGAAGTCTTCGGAAGATATTCGTTTATGATTAACAAT 5946  
QY 1741 gatcaggaagagatgcaaaaaaattccagagctcttgcgaagctactcataaanaatgt 1800  
DB 5947 GATACCGAAGAGATGCAAAAAATTTTCCAGTCTTTCAGGCTTACTCATAAAAATTTGT 6006  
QY 1801 ggagtaaccccaactacgttcttata tgaagcgttgccttcttattcttagctccttcttgc 1860  
DB 6007 GGAGTAACCCCAACCTACGTCCTTATGAAAGCTGCTTCTTATTTCTGAGCTTCCCTTGTTC 6066  
QY 1861 ttaatggcttccctcagctgtagtcttgaactcatcaagaagcgtgtgaa 1907

|||||  
DB 6067 TTAGGGTCTCAGTACGCTTTGACTACATCAAGAGCGTGAA 6113  
RESULT 3  
AE001587/c 16448 bp DNA linear BCF 01-DEC-2000  
DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.  
ACCESSION AE001587 AE001363  
VERSION AE001587.1 GI:4376271  
KEYWORDS  
SOURCE Chlamydia pneumoniae CWL029.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (bases 1 to 16448)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W.,  
Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 2 (bases 1 to 16448)  
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,  
Grimwood, J., Davis, R. W. and Stephens, R. S.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University  
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
location/Qualifiers  
FEATURES  
source 1..16448  
/organism="Chlamydia pneumoniae CWL029"  
/strain="CWL029"  
/db\_xref="taxon:115713"  
misc\_feature 172..2957  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
gene 172..705  
/gene="pmp\_3.1"  
/note="Cp0014"  
615..2957  
/gene="pmp\_3.2"  
/note="Cp0015"  
3196..4971  
/gene="pmp\_4.1"  
/note="Cp0016"  
3196..5953  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
gene 4874..5953  
/gene="pmp\_4.2"  
/note="Cp0017"  
6302..7786  
/gene="pmp\_5.1"  
/note="Cp0018"  
6302..9139  
/note="Polymorphic Outer Membrane Protein G Family;  
contains frame shift"  
gene 7787..9139  
/gene="pmp\_5.2"  
/note="Cp0019"  
complement(9380..11467)  
/gene="Cp0020"  
complement(9380..11467)  
/gene="Cp0020"  
/codon\_start=1  
/transl\_table=1  
/product="Predicted OMP [leader peptide: outer membrane]"  
/protein\_id="AADI8173.1"  
/db\_xref="GI:4376272"  
translation="MKRCFLFIASFVLMSSADALTHQPAVKKKNSYLSHPKSVGIV  
TIDGVNTHNNLRIGQANKYVENTVQSLKVAHGVMYNYRAKTIYCYLEYEET  
DSCLLNGRFAMTFLGSSMITLTPEPTIVIRKQISTSEPKKDLCSGLYLEYSSD  
SLSLGKTTLRVCRIPILFLPESIMPEIPEKPIINRGGGGLGSLGSSYPIR  
KHFSFPLDSFFKHGVMGFNLHCSOKVPENFNMKSYAHLAIIDMAEADHRYRL

HDGCEFTHKHVNESGEYHLSDSMETVADIPNNFMKNTGPTRVCTQWANDYEGYLT  
SSYKVNFEQANQDELPLYTLROYDISLYNGVYLENVEGYNLFAFSDHIVENSS  
LRLAARPKLKVPLPIGTLSTSGSLIYVSDPELSSQSLAKLOAFENLSS  
SYIORRHIIPEPVTEITETREPLAKNEPDHYIETSDAPHSNLLKAGDTSVSKTNR  
PPRIHAKLWTTHTLSTESKPTPEKTRACELSLPECKKNTYSLAEWITMKHCHDHANI  
RWEMIGDNVAMTLESLSHRKSYSLIKCDRENFLIDVSRPIDQLDSPLSHRNLILGK  
LVPVPHPCNNYRLSLRYGWHRODTPNLEYQMILGTIKIFPHWOLYGVYERREADSRF  
FPLKIDPKKPPF"  
complement(11484..13190)  
/gene="Cp0021"  
complement(11484..13190)  
/gene="Cp0021"  
/codon\_start=1  
/transl\_table=1  
/product="Predicted OMP [leader peptide]"  
/protein\_id="AADI8174.1"  
/db\_xref="GI:4376273"  
/translation="MGLFHLTLFGLCLSLPISLVAKFPESVGHKILYSTQOAL  
ATYLEADAYGDHDFVLRKIGEDYLSKOSIHSPPQPKSTIIGAGLAGSSALDVS  
QAMETADPLQOLVLVSGSLGTSDDLFLKALASPYPIVLEAAYRLANLKTQYI  
DHLSFTHLKPEETIOCLSAIPLFLETESDAYIRDLAKKSAIRSATVLOIGEXO  
KRLPTIRNLITSASPODORAILIYALGKLKDGOSYNIKQLOKPDVDTLAAQOAI  
ALGKEEDALVIKQALEERPRALIRHPSEIGIPIALPIFLKTNSSAKLVAILA  
LLEIGCDPTKLELYTERLVQPHNETLALFSKGRQLQMKRKNVILIPODDERKL  
LSTTRGLEQILTEFLRPLPKYALPCYKLLASQKTLATTAISFLTSHQALDL  
FOAKLEGEPIIRAYADLAIYNLKDEKRSKLDHYAKKLIQETILFVDTENGPHRS  
MPYIRYGVTPESRKLMDLLETATLSKSSBDIRLLIQLTMEGDAKNFVLAGLLIKI  
VE"  
complement(13172..13762)  
/gene="maf"  
/note="Cp0022"  
complement(13172..13762)  
/gene="maf"  
/codon\_start=1  
/transl\_table=1  
/product="Mat-type protein"  
/protein\_id="AADI8175.1"  
/db\_xref="GI:4376274"  
/translation="MSPLVIGSSSPRRKFLKEKRVPEFYIPSPESKVSYSGDPI  
AYQELAAOKAYAYSELSHPCDCLIGDPTVSGDGLFTRPODKADAIOMLKTLRNO  
THDVYSIVAVLHKGLTGSETSOISLTMIPDHIESYIDVGTLLNCGAIDVCHGL  
ILKRVHGCYVNVQGLPIQTLKYLEELNIDLMVSI"  
complement(13791..15383)  
/gene="Yj1k"  
/note="Cp0023"  
complement(13791..15383)  
/gene="Yj1k"  
/codon\_start=1  
/transl\_table=1  
/product="ABC Transporter Protein AtPase"  
/protein\_id="AADI8176.1"  
/db\_xref="GI:4376275"  
/translation="MSIVLDIGKSLGTRILFDVSYVENGNCGYLGTPGACKSTL  
LKITGMIEPTRGISLSPKVGILRQNIIDSHDTVLDVCMGTRLEALQORDNLY  
LOEFTDAIGMELGIEIEIGENGYRADSSEBELTIGIPNEFDRKMANIPIDLF  
RVLLQALGFHPEALILDEPTNHDVLSIMLGNFLKDYEGTVVSHDRFLITIT  
HIADIDVTIITYGNVDMEMKTAAREOKADIKSEKIKISOLKEVAFAGWSRA  
SOVSRLREIKLOPQELKKSNIORPIRPLSPKSGKVVLSLEAITKQVGDQOYH  
PFSLEITQGDRLGITGNNGLGKTLTKMLACVEAPSSGKILGHQALCSITPQNHSDV  
LADGQETLFEMLNRKTGINDQIRSVLGMELGGDAFQIQALSGGETARLLMAG  
MLENHNVLILDEANNHLDLESVSALSWAINDYGTAFIVSHRDLIODCATKLIIFD  
KDKITFPDGTVDYTAGHKOLL"  
complement(15438..16376)  
/gene="xerc"  
/note="Cp0024"  
complement(15438..16376)  
/gene="xerc"  
/codon\_start=1  
/transl\_table=1  
/product="Integrase/recombinase"  
/protein\_id="AADI8177.1"  
/db\_xref="GI:4376276"  
/translation="MISIVSFLDYLKMKVKSASPHTLRNVCCLDNLGLKIFLERGNIA

PSPLQATLTERKRVSELPESLPTKEHVAMYLAKLIENKARTIKRCLSSIKSPAHVC  
VIOKILENPEPETHGPRKELPSPMYIAQVEVLMATPDLSKTHGLDRLMELFYS  
SGNLSIENVANKODFDLSTHLIRGKAKERIIPVSNAMOWQIYLNHPRKRE  
KDPQALFLNRGRIRISIRIDRSFOEYLRJRSGLSHITPHIRHTIAHWLESGMDLK  
TIOALFLGSSLETTVYTOVSVKLKQTHQAHBPHA"  
BASE COUNT 4710 a 3330 c 3665 g 4743 t  
ORIGIN

Query Match 100.0%; Score 1907; DB 1; Length 16448;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1907; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtggttgatttgaagaaagtccatgagtgtttaaagttcaaggtccctacc 60  
Db 13290 GTGGCTTATTTGAAAAAGGCCATGATGCTTATTAATGAAGTCCCTATACC 13231  
Qy 61 aaacatgaaatctgtctagagaggttgaacatgacatagggactatccatcac 120  
Db 13230 AAACATGAATACTTGTCTAGAGAGCTTGAACATGATCTATGGACTATTCATCTAAC 13171  
Qy 121 tctcttggaactttatgtgtgagctctccatctctctgtctgtaaatccctgagtc 180  
Db 13170 TCTCTTGGACTTTATTTGTGAGTCTTCCATTTCTCTTGTGCTAATTCCTGAGTCC 13111  
Qy 181 tgtaggtcataagatccctttatataagtaagcaatctcaacagagccttagaacaata 240  
Db 13110 TGTAGGTATTAAGATCCTTTTATTAAGTACCAATCTACACAGCGGCTTGAACAATA 13051  
Qy 241 tctggaagccttaagatgctcagctgcatgacatctctgtctttaaagaaatctggaga 300  
Db 13050 TCTGGAACCTTACATGCTGCTACGGGATCATGACTTCTTCGTTTAAAGAAAATGGAGA 12991  
Qy 301 agactatctcaagcaaaagcatccatctcctcagatccgcaaaactagaagaaagcaccatcat 360  
Db 12990 AGACTATCTCAAGCAAAACATCCACTCCTCAGATCCGCAACTAGAAAAAGCACCATCAT 12931  
Qy 361 tggaggaagccttgaggatcttcaagaagccttggagctgtctcccaagctatgaaac 420  
Db 12930 TGGAGCAAGCCTGGGGATCTTCAGAAAGCTTGGACGTGCTCCCAAGCATATGGAAC 12871  
Qy 421 tgcgaacccctgcagagcactggttttatacggcagctcccaagacatcttggagaaac 480  
Db 12870 TGCAGACCCCTGCAGCAGCTACTGTTTATCGGCACTCTCAGAGACATCTTGGAAAAAC 12811  
Qy 481 tctggaagcactggtttaaagctttagacatccctcctcgtlcatccgcttagaagc 540  
Db 12810 TTTCTGACACTTACTGTTTAAAGCTTTAGCATTCCTCATCTCCTTGAGAACG 12751  
Qy 541 cgccatagacttgctaatttgaagaacactaaagctatgatactcaatcttcttcat 600  
Db 12750 CGCCTATAGACTTGCTAATTTGAAGAACATTAAGTCAATGATCATCAATCTTTTAT 12691  
Qy 601 tcatatagcttccgaaagaatccaatgcttacttgcggaacaatctcagccttggagac 660  
Db 12690 TCAATAGCTTCCGAGAAATCCAAATGCTATCTGCGGCAATATTTCCATCCCTTGGAGAC 12631  
Qy 661 tgaagaatctgagtgctatatactcgggagctcttagctgccaagaagcgcgattcgag 720  
Db 12630 TGAAGAATCTGATCTTATTTATTCGSGATCTCTTAGCTGCCAAGAAACGCGATTGCGAG 12571  
Qy 721 tgcacagacttgagatcggaagaatccaaacaaagccttctctccgaacaattagaa 780  
Db 12570 TGCACAGACTTTCAGATCGGAGATACCAACAAAAGCTTTCTTCGCAACTTAGAGAA 12511  
Qy 781 ttgtctaagagtgcttctcccaagatcaagaagctatctctttagcttgaaggaagct 840  
Db 12510 TTGCTTAACGAGTGTCTCTCTCTCAAGATCAAGAGCTATTTCTTATGTTTAGGGAAGCT 12451  
Qy 841 taaggaatgctcagagctactacaatataaaaaagcaattgcagaagcctgagtgtgagt 900  
Db 12450 TAAAGATGCTCAGAGCTACTCATATATAAAAAAGCAATTGCAGAAAGCCTGATGTGAGATG 12391

Qy 901 caacttagacagagctcaagcctttaaattgcttctgggaaagaagagcagctctccgct 960  
Db 12390 CACTTAGCAGCAGAGCTCAAGCTTATTTCTTTGGGGAAGAAAGAGAGCTCTTCCCT 12331  
Qy 961 gataaaaaagcaagcacttgagagagcctcgaagccctgtatgcttaagcattacc 1020  
Db 12330 GATAAAAAGCAAGCCTTGAAGAGGCGGCTGAGCGCTGTATGCTTACGCTATTCAC 12271  
Qy 1021 ccttgagaatgagatctccgacttgccttcggaatattcctaaanaactaaagaacggaagc 1080  
Db 12270 CTCTGAGATAGGAGATTCGATTCGCCCTGCCGATTTCCATAAAACATAAGACGAGAAC 12211  
Qy 1081 caagttgaatgagctttagctctctcttagagtgagtgagtgacacccctaaactactgga 1140  
Db 12210 CAAGTTGAATGTAGCTTTAGCTCTCTTATAGATTAGGGTGTGACACCCCTTAACCTACTGGA 12151  
Qy 1141 atacattaccagaagagctgtgtccaaacaaatataatagagactcragccttgaagttctc 1200  
Db 12150 ATACATTACCGAAGAGCTTGTCCAAACCATTAATTAAGACTCTAGCCTTAGATTCTC 12091  
Qy 1201 taagggcgctacttacaanaattggaaagcggtgaaacatcatagctccctcaagaatcccca 1260  
Db 12090 TAAGGGCGCTACTTTACAAAATTTGGAGGGGTGAACATCATAGTCCCTCAAGATGCCCA 12031  
Qy 1261 ggaagaggaaaggttgctctccacaacccgaggtcttgaagagcagatcccttagttctc 1320  
Db 12030 GGAGAGGGGAAGGTTGCTCTCCACAACCCGAGGCTTTGAAGAGCAGATCCTTACGTTTCT 11971  
Qy 1321 ctccgctacccttaagaagagcttaccctccctgtattttaaagccttggggaagtcaga 1380  
Db 11970 CTTCGGCTACTTAAGAAAGCTTACCTCCCTGTAATTTATTAAGCTTTTGGCAGTCCAGAA 11911  
Qy 1381 aactcaagcttgcacactgagcttcttctttaaagtcacacactacatcaagaagcctt 1440  
Db 11910 AACTCAGCTTGCCACTACTGCGATTTCTTTTATAGTCAACACTCAATCAGAGAGCGCTT 11851  
Qy 1441 agatctacttctcagaagcttggaagccttccctggaaacatcatcccgagcctatgacga 1500  
Db 11850 AGATCTACTTTTCCAAAGCTGCGAAGCTTCCGGAACCTTCAATCAATCCGCGCTATGACGA 11791  
Qy 1501 tcttgctatttaacttccaccaaagaatccctgaagaaacagctctctccatgattatgc 1560  
Db 11790 TCTTGCTATTTATTAATCTCACCAAGATCTGAAAAAAGCTTCTCTCATGATTAATGTC 11731  
Qy 1561 aaaaaagctaatctcagaaaccttgatttctgtgacacggaagaaaccaaagaccccatcc 1620  
Db 11730 AAAAAAGCTAATTCAGGAACCTTGTATTGTGAGACAGGAAACCAAAAGACCCCATCC 11671  
Qy 1621 cagcagctctatctagcttataagtgtaaccccgagaagcgttaagaagctcatgttga 1680  
Db 11670 CAGCATGCGCTTATCTTACGTTTATCAAGGTACCCCAAGACCGTACGAAGCTCATGTTGGA 11611  
Qy 1681 tatctagaagaacttagcagacccctcgaagctcttcgaaagataatcgttattgatcaact 1740  
Db 11610 TATTTCTAAGACACTAGGCACCTCGAAGCTTTCGAAAGATATCCCTTATTTATGATCAACT 11551  
Qy 1741 gatgacggaagagatgcaaaaaaattccagctctctcgaagcttactcatalaanaaattgt 1800  
Db 11550 GATGACGGAAGAGATGCAAAAAAATTTCCAGTCTCTTCAGAGCTTACATCAATAAATTGT 11491  
Qy 1801 ggaagtaaccccaactaagcttataagaaggttcttcttatttcttaagcttcccttgttc 1860  
Db 11490 GGAGTAACCCCAACCTTAGCTTATGAAGCGTGGCTTCTTATTTCTTACCTTCTTGTTGC 11431  
Qy 1861 ttatgggttctcagctgagctttagactcaatcaagaagcgtgtgaa 1907  
Db 11430 TTAATGGTCTCTCAGCTATGCTTTGACTCATCAAGAGGCTGTGAAA 11384

RESULT 4  
AX349527  
LOCUS AX349527 1707 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 50 from Patent WO0202606.

ACCESSION AX349527  
VERSION AX349527.1 GI:18615370  
KEYWORDS Chlamydia pneumoniae.  
SOURCE Chlamydia pneumoniae.  
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
REFERENCE 1 (sites)  
AUTHORS Ratti, G. and Grandi, G.  
TITLE Immunisation against Chlamydia pneumoniae  
JOURNAL Patent: WO 0202606-A 50 10-JAN-2002;  
Chiron S.P.A. (IT)  
FEATURES  
source Location/Qualifiers  
1..1707  
/organism="Chlamydia pneumoniae"  
/db\_xref="taxon:83558"  
BASE COUNT 493 a 415 c 340 g 459 t  
ORIGIN  
Query Match 89.5%; Score 1707; DB 6; Length 1707;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1707; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 101 atgggaacttccatctaacctctcttggaacttatttgtagtcttcccatcttctt 160  
Db 1 ATGGGACTATTCATCTCACTCTTTGGACATTATGTGTAGTCTCCATTTCTCTT 60  
QY 161 gtgtgcaaatctccgtgaactctgttagtcttaagatcccttataaagtagcaatctaa 220  
Db 61 GTTGCTAAATTCCTGAGTCTGTAGTCAATAGATCCTTATATATAGTACGCAATCTAC 120  
QY 221 cagcagagccttagcaacatctctggaagctctagatgcctacggtatcatgactcttc 280  
Db 121 CAGCAGGCTTGGACATATCTGGAAGCTTAGATGCTACGATCATGACTCTTTC 180  
QY 281 gtcttaagaaaaatcggaagaactctctcaagcaaaagcatccactccctgaatccgcaa 340  
Db 181 GTTTTAAGAAAAATCGAAGACATCTCAAGCAAAAGCATCCACTCTCAGATCCGAA 240  
QY 341 actaagaaaaaagcaactctgtgagcagcctgagcgtcttcaaaagccttgagcgtg 400  
Db 241 ACTAGAAAAAGCACATCTGTGAGGAGCGCTGGCGGATCTTCAAGCGTTGGACGTG 300  
QY 401 ctctcccaagctatggaactgagacccctgcaagcagctactggtttatcggcagtc 460  
Db 301 CTCTCCCAAGCTATGGAACCTCGAGACCCCTGACAGCTACTGGTTTATGCGAGTC 360  
QY 461 tcaagacatcttgggaaaaactctgaagactactgtttaaagctttagcatctccat 520  
Db 361 TCAGGACATCTTGGGAAAACTTCTGACGACTTACTGTTAAAGCTTTAGCATCTCCCTAT 420  
QY 521 cctgtatccgcttagaagccgctctatagactgtcaatttgaagaacataaagtcatt 580  
Db 421 CCGTGTATCCGCTTGAACCCGCTTATAGACTTGTGAATTTGAAGAACATTAAGCATTT 480  
QY 581 gatcatcatctctctctcaataagctcccgaaagaatccaatgacctatctgggca 640  
Db 481 GATCATCTCATCTTCTTCAATTAAGCTTCCGGAAGAAATCCAAAGCCATATGCGGCA 540  
QY 641 atattctcaagcttggagactgaagaatctgactatatttcgggactcttagctgc 700  
Db 541 AATATCTCAAGCTTGGAGACTGMAAATCGATGATTATTTGGGATCTCTTAGCGGCC 600  
QY 701 aagaagaagcgagcttgaggtgcacagcttgcagatcgggagaatacaacaacaaagc 760  
Db 601 AAGAAAAAGCGGATTCGGAGTGCACAGCTTTCAGATGGGAGAAATCAACAAAAAGCG 660  
QY 761 ttctctcgaacttaggaattgtcaacagtgctctccctcaagaatcaagaagctatt 820  
Db 661 TTCTCTCGACACTTAGGAATTTGCTACGAGTGGCTCTCCTCAAGATCAAGAAGCTATT 720  
QY 821 cttatgctttaggaagcttaagatggtcagagctactacaataaanaagcaatg 880  
|||||

Db 721 CTTTATGCTTTAGGAGCTTAAGGATGCTCAGAGCTACTACATATATAAAAAACAATTG 780  
QY 881 cagaagcctgattgtagtgcactttagcagcagcccaagcttaattgcttgggga 940  
Db 781 CAGAAGCTGATGTGATGTGACTTATAGCAGCGCTCAAGCTTAAATGCTTTGGGGAAA 840  
QY 941 gaaaggaagccttccgttgataaaaaagcaagcaacttggagagcgccctcgaagctg 1000  
Db 841 GAAGAGAGAGCTTCCCGGATAAAAAAGCAAGCACTTAGAGAGCGGCTCGAGCCCTG 960  
QY 1001 tatgctttagcagcatctaccctctggaatagggaatccgattgcttgcggaattccca 1060  
Db 901 TATGCTTAGCGGATCTACCTCTGTGATAGGATTCGATTCCGATTCGATTTCTTA 960  
QY 1061 aaacaaagaagcagcgaagcgaagtgtgaatgttagctttagctctcttaaggtta 1120  
Db 961 AAACCTAAGAACAGCGAAGCAAGTTGAATGTAGCTTTAGCTCTCTTAGAGTTAGGCTGT 1020  
QY 1121 gacacccctaaactactggaatacatcaccgaagaagctgttccaacacataatag 1180  
Db 1021 GACACCCCTAAACTACTGGAATACATTACGGAAGGCTTGTCCACACATTTAATAGAG 1080  
QY 1181 actctagccttgagttctcttaagggcgtacttacaanaattggaagcggtgacatc 1240  
Db 1081 ACTCTAGCCCTTGAAGTTCTCTTAAGGGCGCTACTTACAAAATTGGAAAGCGGTGAACATC 1140  
QY 1241 atagtcacctaaagatcccaagaggaagaagttgctctcacaacccgaagcttgaa 1300  
Db 1141 ATAGTCCCTTAACATCCCGAGAGAGGAAAGTTGCTCTCCACACCGCAGGCTTTGAA 1200  
QY 1301 gagcaagctcttactgcttctctccgactcaactaaagaagcttactccctgtattat 1360  
Db 1201 GACGAGATCTCTTACGTTTCTCTCCGCTTACCTAAAGAACTTACCTCCCTGATTTAT 1260  
QY 1361 aagctttggcgaatgaagaactcagctgcccactactgagattcttlttaagtca 1420  
Db 1261 AACCTTTGGCGAGTCAAGAAACTCAGCTTCCACTGCACTGCAATTTCTTTTAAAGTCAC 1320  
QY 1421 acctacatcagaagaacttagatctacttllcaagctctgcaagctctcttgagaacct 1480  
Db 1321 ACTCATATAGGAAGCTTAGATCTACTTCTTCAMAGCTCGAAGCTTCTTGAGAACCT 1380  
QY 1481 atcatcgcgcctatcagatctctgtatllataatctcacaagaagctctgaaaaaaa 1540  
Db 1381 ATCATCGCGCCCTATGAGATCTTGCTATTATTAATCTACCAAGATCTCGAAAAAAA 1440  
QY 1541 cgttctctcactatgatacaaaaaagctaaatcgaagaaccttgtatttggacag 1600  
Db 1441 CGTTCTCTCATGATTATGCAAAAAAGCTAATTCAGGAACCTTGTTATTTGTGACACG 1500  
QY 1601 gaaaccaaagaagcccatcccaagcctatcagcttactaggtcaccccaagaagc 1660  
Db 1501 GAAACCAAGAGCCCATCCACATGCCCCTATCTACGTTATGAGGTACCCCAAGAAAGC 1560  
QY 1661 cgtcaagaagctcagtgtgatatcttagagacactgcaacccctcgaagctctcgaagat 1720  
Db 1561 CGTACGAGCTCATGTTGTTGATTTCTTAGAGACACTGACCCGCAAGTCTTCGGAAGAT 1620  
QY 1721 atccgcttactgatacaactgtagcaggaagaagatgcaaaaaattccagctcttga 1780  
Db 1621 ATCCGTTTATTTGATACAACTGATGACGGAAGGAGATGCAAAAAATTTCCAGTCTTGCA 1680  
QY 1781 ggtctactactaaaaattgtagagtaa 1807  
Db 1681 GGCTTACTCATTAATAAATTGTGAGATA 1707  
RESULT 5  
AX349569  
LOCUS AX349569 2088 bp DNA linear PAT 06-FEB-2002  
DEFINITION Sequence 92 from Patent WO0202606.  
ACCESSION AX349569  
VERSION AX349569.1 GI:18615391



```

KEYWORDS
SOURCE      Chlamydomophila pneumoniae.
ORGANISM    Chlamydomophila pneumoniae
REFERENCE    Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
AUTHORS     1 (sites)
TITLE       Ratti,G. and Grandi,G.
JOURNAL     Immunisation against Chlamydia pneumoniae
            Patent: WO 0202506-A 92 10-JAN-2002;
            Chiron S.p.A. (IT)
FEATURES
source      Location/Qualifiers
            1..2088
            /organism="Chlamydomophila pneumoniae"
            /db_xref="taxon:83558"
BASE COUNT  574 a      480 c      417 g      617 t
ORIGIN

Query Match      4.4%: Score 84; DB 6; Length 2088;
Best Local Similarity 100.0%; Pred. No. 2.3e-37;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1824 atgaacgttgctctctatctctagctcctctgtctctatgaggtctctcagctatgct 1883
|||||
Db 1 ATGAACGTTGCTCTTATTCTTAGCTTCCTTTGTTCTTAIGGTTCTCAGCTATGCT 60
|||||

QY 1884 ttgactcatcaagagagctgtgaaa 1907
|||||
Db 61 TTGACTCATCAAGAGCGCTGTGAAA 84
|||||

RESULT 6
AX100531.c 36 bp DNA linear PAT 10-APR-2001
LOCUS
DEFINITION Sequence 4 from Patent WO0121804.
ACCESSION AX100531
VERSION AX100531.1 GI:13619535
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
REFERENCE
AUTHORS Mordin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE I chlamydia /I antigens and corresponding dna fragments and uses
        thereof
JOURNAL Patent: WO 0121804-A 4 29-MAR-2001;
        Aventis Pasteur MSD (FR)
FEATURES
source      Location/Qualifiers
            1..36
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="3' PCR primer"
BASE COUNT  8 a      12 c      7 g      9 t
ORIGIN

Query Match      1.3%: Score 24; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.066;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1781 ggcttactcataaaattgtggag 1804
|||||
Db 36 GGCTTACTCATTAATAAATTGTGGAG 13
|||||

RESULT 7
AX100530 44 bp DNA linear PAT 10-APR-2001
LOCUS
DEFINITION Sequence 3 from Patent WO0121804.
ACCESSION AX100530
VERSION AX100530.1 GI:13619534
KEYWORDS
SOURCE
synthetic construct.
ORGANISM
synthetic construct.

```

REFERENCE	artificial sequence.			
AUTHORS	1 (bases 1 to 44)			
TITLE	Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.			
JOURNAL	I chlamydia /1 antigens and corresponding dna fragments and uses thereof			
FEATURES	Patent: WO 0121804-A 3 29-MAR-2001;			
source	Aventis Pasteur MSD (FR)			
	.location/Qualifiers			
	1..44			
BASE COUNT	12 a	14 c	8 g	10 t
ORIGIN				
Query Match	1.3%	Score 24;	DB 6;	Length 44;
Best Local Similarity	100.0%	Pred. No. 0.066;		
Matches 24;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy 101	atggagactatccatctactctc 124			
Db 21	ATGGAGCTATTCCATCTACTCTC 44			
RESULT 8				
AE001308				
LOCUS	AE001308 11627 bp DNA linear BCT 30-OCR-2000			
DEFINITION	Chlamydia trachomatis section 35 of 87 of the complete genome.			
ACCESSION	AE001308 AE001273			
VERSION	AE001308.1 GI:3328766			
KEYWORDS				
SOURCE	Chlamydia trachomatis.			
ORGANISM	Chlamydia trachomatis			
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
AUTHORS	1 (bases 1 to 11627)			
	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,			
	Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,			
	Koonin,E.V. and Davis,R.W.			
TITLE	Genome sequence of an obligate intracellular pathogen of humans:			
JOURNAL	Chlamydia trachomatis			
MEDLINE	Science 282 (5389), 754-759 (1998)			
PUBMED	99000809			
REFERENCE	9784136			
AUTHORS	2 (bases 1 to 11627)			
	Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,			
	Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.			
TITLE	Comparative genomes of Chlamydia pneumoniae and C. trachomatis			
JOURNAL	Nat. Genet. 21 (4), 385-389 (1999)			
MEDLINE	99206606			
PUBMED	10192388			
REFERENCE				
AUTHORS	3 (bases 1 to 11627)			
	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,			
	Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,			
	Koonin,E.V. and Davis,R.W.			
TITLE	Direct Submussion			
JOURNAL	Submitted (20-MAY-1998) Program in Infectious Diseases, University			
FEATURES	of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA			
source	Location/Qualifiers			
	1..11627			
	/organism="Chlamydia trachomatis"			
	/strain="D/UW-3/CX"			
	/db_xref="taxon:813"			
	207..1121			
	/gene="elac"			
	/note="CT346"			
	207..1121			
	/gene="elac"			
	/codon_start=1			
	/transl_table=11			
	/product="Sulphohydrolase/Glycosylatase"			
	/protein_id="AAC67941.1"			
	/db_xref="GI:3328767"			
CDS				
gene				



gene  
1183. .2130  
/gene="xerc"  
/note="CT347"  
1183. .2130  
/gene="xerc"  
/codon\_start=1  
/transl\_table=11  
/product="Integrase/recombinase"  
/protein\_id="AAC67942.1"  
/db\_xref="GI:3328768"  
/translation="MITSFYAFDLYLKNKASLTLRNYCDLSLCKFLKKSDLS  
PTPLSLDNTYDPLPSFLTKDNIRLYLEQIQTHSKRTVRRISAIKSFARFC  
VKNQLIPENPAMIRGPRLOELPSPLYEOVLALMAPELDKVTGFRDCLLELYS  
SOLRISEITALNRADIDFOSHLHIRGKKERIVPMTRVAQOMLODYLNHDDRAVE  
ODHOACFLNRFGRRLSTRSIDRKFOOYLTKGTSGSIPPTIRHTIATHWLERGMDLK  
TIOLLGHTSLETTTIYTHVSMKKLKKQIHDEHPHNLPE"  
2184. .3770  
/gene="yjjk"  
/note="CT348"  
2184. .3770  
/gene="yjjk"  
/codon\_start=1  
/transl\_table=11  
/product="ABC transporter Protein ATPase"  
/protein\_id="AAC67943.1"  
/db\_xref="GI:3328769"  
/translation="MSIVLDKIGKTLGTPRVLPDVSYPENPGNRYGLNGPAGKSTL  
LKIIGIVPESRGTISLPKIGILRONIDSGDVIIVDCVTMGSRIMAMGRBALY  
AEFFDAIGMELGEIEBITIGEDGYRAESEAEELLIGIPEEPFSQKNATTPLDJQF  
RYLLCQALGHPHEALLDEPTNHLDLHSLNMGSLFKDYNGTIVVSHDRHFLNTTT  
HLADIDYDTIITPGNDAMVAMKASRQEKADIKSEKKTAQLREFYAKGASRA  
SOVORLREIKIQPOELKKSNIQRPYIRFVPSDKASGVPSIGISKTYPADPLFK  
PSLEIYOGDKIGVINGNGLTGTLMLKLAGVEYDSOGQVKGHNVAEYFPONADI  
LKDCGEITLFEWLRRRKGTIPDOEYRSVIGKMLRGDDAFKQVKALSGETRRLLMAG  
KMLENNHMLIDEANNHLDLESVSALAMAINDPGATLFSHRTLIDECATRLILFN  
KGTISFGGPMSEYLAOSK"  
3806. .4396  
/gene="maf"  
/note="CT349"  
3806. .4396  
/gene="maf"  
/codon\_start=1  
/transl\_table=11  
/product="Maf-type protein"  
/protein\_id="AAC67944.1"  
/db\_xref="GI:3328770"  
/translation="MEARLVLGSSSESRKAVLESRRIPICVSPDFEDRSIYSGDP  
EYTKELAMKANAVVSQGSFALITADIVVYKGEVFKPSEEHAVMLPTLSGSS  
HSVITTLVLMONEKVLASSENTQVSFIDIPQHLTLTYVRSFSLKRCGGYCVQDGGGL  
ITKIQEGCVYNIQGLPIKTLNQLMELFNIDYLV"  
4378. .6078  
/gene="CT350"  
4378. .6078  
/gene="CT350"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC67945.1"  
/db\_xref="GI:3328771"  
/translation="WGLSRLATISPLSFTLSASCDPSPSSVSGRIELSCRSVQAIEA  
YLEASATYQOHDFSVLRVLAESYLOQSFLESDTYIRKSAIIGAGLSGSEALELSEA  
IETQDLYEQLILINATSQLSKTSKTLFGKLTGASHPYIRLEAAVYLAGMKASLSDY  
LVSFLYKLPBEEIOLAAATFLOLTFEADAYIHLHLSPPNNLIRNVAVVLAIGEYKQR  
FLPTLRSLITSASPLDEGALAIKGLDNGSYPRIKALSRSNPENVVLAAGOTILFL  
EKEERALPITNLCQOKILRAIYTARFISQEKGBELLPIFYNATODELRINTALAV  
HOGCTDPOVLAHTLEILSKVLHRIPLPHSTGKAIQFWKECTFPLMSQDEKMRILA  
MYVAEDTILSALKIIPNDAYLPIERITLASQKITLAANAIAFLSVTAHPQALSTVSK

gene  
6085. .8178  
/gene="CT351"  
6085. .8178  
/gene="CT351"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC67946.1"  
/db\_xref="GI:3328772"  
/translation="MKREFPIGVLAHTLPSEGLSHQQAQVKISYLSHPKGTGI  
MDEVGLVHIHDDLRLQANKAVENRTDGIKIVAGNVMVYRGKILICDYLEVED  
TDSCLTNGRCGLYPMFTGGSTITISPSIILHKGIISSEGPKHILCSDYLKXSS  
DSVLISMGPSRLSICMPVILLPOISIMEPEKPIPTFRSGSGGLGSLGYSPIS  
KHCSTTLPDGFERNHGIGLGNMFSSSEENSNAINIKSYAHLRAIDSSAKDRIR  
LHGDFTFESKERAHLAGEPILSDSWETVVDIFPNMFSLKATGPEVLSLRDNNLFKM  
TSSVYNSFQNVKQELPOLHHRVRIARRRIPIENRLEAGFLDFHFNSSNIPGSNFS  
SWRFSNAHVYGLVPIGTLTSPSLSGTAIYTYRMLSNPAACOLSGSLSPDYRYALQ  
KEYRAHRHIVEPFGSLKTRVLSDEPHIISIDAFHSINLHVGLSKYLVNHSI  
PSHLKIMTYIIRDEPHAKDTPKTAQMSFPLTQNTSISDAETWKSRRMHLNVIY  
EWILNDNLGTLTEFLHRSKYGTIKCAKNDYTLVDQSGSLDTLASPLSDRRNITGKLF  
VRRHPHMNYNLMRLRYGWARPDSPSYLEYQMLIGHKIFEHMQLFSYEKREADKRCFFY  
LKLDRKKOKHRHPCG"  
8341. .8646  
/gene="CT352"  
8341. .8646  
/gene="CT352"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC67947.1"  
/db\_xref="GI:3328773"  
/translation="MRLNLIKKKIQLTRLPFSKFLGILIOINLRNRHKKTLTIPC  
SYSLQEQVHTLSITKGTIRTKCKTEQPPYPKILGHTRKLLALNPTFICHSAV"  
complement(8716. .9261)  
/gene="def"  
/note="CT353"  
complement(8716. .9261)  
/gene="def"  
/codon\_start=1  
/transl\_table=11  
/product="Polypeptide Deformylase"  
/protein\_id="AAC67948.1"  
/db\_xref="GI:3328774"  
/translation="MIRDELYDSPLRKVAVPEITDELQQLVLDMEFMATFKGV  
GLAAPQVGQISLFTMGVRELEDELVECDPRVFINFVITQKSOALVMEGECIST  
PGLRGEVARPDKITVSAKNLDGQQFSLALEGLARIYMHETDHLGVLYIDRMSDK  
TKQFKNNLEKIRRKYSILRGL"  
complement(9536. .10369)  
/gene="kgsa"  
/note="CT354"  
complement(9536. .10369)  
/gene="kgsa"  
/codon\_start=1  
/transl\_table=11  
/product="dimethyladenosine Transferase"  
/protein\_id="AAC67949.1"  
/db\_xref="GI:3328775"  
/translation="MARSSIEQLTFSLRVNGRAKKALSONFLVDGNILTKLITTAEV  
OPGDWVLEIGPGFALSEVLISQGANVIALEDPMFEESLSQLPMDIETTDACKYPLT  
SLDGKGMKGKGRIVANLPHYITTPLTCKFLPCQPYKWKIVTVMIDQEVARITAKGD  
KXYSITVTSFPAVOYAFKYSNCFNCPYKSHVSNVYMRVHEDQALADSIIEEFT  
LRRAFGQRKRLIANSLKMLYPRKDYFOVLQDLSGSEKTRPETIFLEETLIKPHLLKD  
I"  
complement(10385. .11446)  
/gene="CT355"  
complement(10385. .11446)  
/gene="CT355"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"

```

/protein_id="AAC67950.1"
/db_xref="GI:3328776"
/translation="MRVERAVIVACIYITIFSGIARCYEGACSSGSEPNQSCVAIH
NRVLFKDEDEVTTLDTIRKLNILFISTCPOLVDVSARSQYVSAMPVLETVINE
FLMADAKAKKIFIDPVSNOEIAMFGERDLSPPAKFEDMDPEVNVNHLIAQRY
EGMVRSRVMLKTPGMVREYQKLADEAQTOMTYKVLTKAGLEFLAKIKGVO
ERLNEGSSMDKERLTALVLSOGOLMCSEEPFRDAOLSVNHRQSLSEINPERCGK
VLEHAGSLKLEVLNRAIKTLDPDKMEAOIKOOLMEFAAEENANKKRLHARYGPD

Query Match      1.2%: Score 22; DB 1; Length 11627;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 519 atccgtcatccgcttagaagc 540
|||||
Db 4790 ATCCTGTCATCCGCTTAGAGC 4811

RESULT 9
AC017438 47445 bp DNA linear HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
DEFINITION pieces.
ACCESSION AC017438
VERSION AC017438.1 GI:553548
KEYWORDS HTG; HTGS; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 47445)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDW:10210763 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..47445
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 14508 a 9495 c 8973 g 14469 t
ORIGIN

Query Match      1.2%: Score 22; DB 2; Length 47445;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 860 tacataataaaagcaattgc 881
|||||
Db 34100 TACAATATAAAAAGCAATTGC 34121

RESULT 10
AC009341 183722 bp DNA linear INV 10-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2L, region 36C-36C, BAC clone
DEFINITION BACR07G13, complete sequence.
ACCESSION AC009341
VERSION AC009341.6 GI:13270516
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 183722)
REFERENCE
1

```

```

AUTHORS
Celisner, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Bazon, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Chame, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnak, D., Farfan, D.,
Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibegwam, C., Jalali, M., Kuse, D., Li, P., Matel, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuno, J.,
Pacled, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirkas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
TITLE Sequencing of Drosophila chromosome 2L, region 36C-36C
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183722)
AUTHORS
Celisner, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Butenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Katra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomont, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Mar 10, 2001 this sequence version replaced gi:7264746.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
FEATURES
source
1..183722
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="36C-36C"
/clone_id="BACR07G13 (D1029)"
/clone_id="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
PBACe3.6)"
BASE COUNT 54821 a 35450 c 37135 g 56316 t
ORIGIN

Query Match      1.2%: Score 22; DB 3; Length 183722;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 860 tacataataaaagcaattgc 881
|||||
Db 50434 TACAATATAAAAAGCAATTGC 50413

RESULT 11
AE003654 261000 bp DNA linear INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386055 section 47
DEFINITION of 63, complete sequence.
ACCESSION AE003654
VERSION AE003654.1 GI:7298369
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```



/db\_xref="FLYBASE:FBan0015139"  
/db\_xref="FLYBASE:FBgn0040987"  
/evidence=not\_experimental  
115662..115853  
/gene="CG15139"  
/note="CG15139 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0015139"  
/db\_xref="FLYBASE:FBgn0040987"  
/evidence=not\_experimental  
/protein\_id="AAF3599.1"  
/db\_xref="GI:7298373"  
/translation="MAVINPQWRPASDLRSQSTIPPKDVNDAAEEQQEQEHIIISGDS  
VSHEDISLVSGLNSPN"  
complement(join(<127732..127918,136653..136803,  
136950..137197,138481..138519))  
/gene="CG6365"  
/product="CT19870"  
/db\_xref="FLYBASE:FBan006365"  
/db\_xref="FLYBASE:FBgn0032630"  
complement(<127732..>138519)  
/gene="CG6365"  
/map="36B5-36B6"  
/db\_xref="FLYBASE:FBan006365"  
/db\_xref="FLYBASE:FBgn0032630"  
complement(join(127808..127918,136653..136803,  
136950..137197,138481..138519))  
/gene="CG6365"  
/note="CG6365 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan006365"  
/db\_xref="FLYBASE:FBgn0032630"  
/protein\_id="AAF3600.1"  
/db\_xref="GI:7298374"  
/translation="MAIASYLGAAPVANKTVKPYKTEVTHRIKRALRLVLSAGIVATS  
VPHREFRLDTKNGKMYLKDTESLFREHNGSOVDWINGVALSVICVCPOTKY  
DRKOLINSKGTAVKONQVIEIKVFNILGILKRRKNQSLSKSKYSISGAGSHKHKSA  
NNOHWPSPQAFSGSVFRRTER"  
join(<162848..162875,162979..>163769)  
/gene="CG15140"  
/product="CT15037"  
/db\_xref="FLYBASE:FBan0015140"  
/db\_xref="FLYBASE:FBgn0032631"  
/evidence=not\_experimental  
<162848..>163765  
/gene="CG15140"  
gene

Query Match 1.2%; Score 22; DB 3; Length 261000;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 860 tacaataataaaagcaattgc 861  
|||||  
Db 41537 TACAATATATAAAGCAATTGC 41516

RESULT 12  
LOCUS AX035433 4343 bp DNA linear PAT 15-NOV-2000  
DEFINITION Sequence 8 from Patent EP1035209.  
ACCESSION AX035433  
VERSION AX035433.1 GI:11191075  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial sequence.  
REFERENCE 1 (bases 1 to 4343)  
AUTHORS Azzey,M., Hobom,G., Wenke,A. and Flick,R.  
TITLE Stable recombinant influenza viruses free of helper viruses  
JOURNAL Patent: EP 1035209-A 8 15-SEP-2000;  
ARTEMIS PHARMACEUTICALS GMBH (DE)  
FEATURES  
source Location/Qualifiers  
1..4343

/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="PH12989"  
BASE COUNT 1085 a 1064 c 1100 g 1094 t  
ORIGIN

Query Match 1.1%; Score 21; DB 6; Length 4343;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1546 tctcatgatattgcaaaaa 1566  
|||||  
Db 1876 TCTCATGATATTGCAAAAA 1896

RESULT 13  
LOCUS AL670958 177014 bp DNA linear HTG 24-JAN-2002  
DEFINITION Mus musculus chromosome 4 clone RP23-410K19, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*. In unordered pieces.  
ACCESSION AL670958  
VERSION AL670958.1 GI:18376488  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_ACTIVEFIN.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 177014)  
CLARKE,D., CONNOR,R., LEAVES,N.I., CAYEBERRY,L., GREYSTONG,J.,  
NORTH,P.C., HUNTER,G., SHUFFLEBOTTOM,L., KIMBERLY,C., CAMPBELL,M.,  
JONES,S., LAWRENCE,N., STRACHAN,G.L., GREENHAM,L., MAGGOTT,K. and  
BOTCHERBY,M.R.M.  
Direct Submission  
Submitted (23-JAN-2002) Mouse Sequencing Group, HGMP-RC, Hinxton,  
Cambridge, CB10 1SP, UK. E-mail enquiries:-- mbotche@hgmprc.ac.uk  
or clare@hgmprc.ac.uk  
HGMP-RC part of the UK Mouse Sequencing Consortium  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved  
Location/Qualifiers  
1..177014  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-410K19"  
/clone\_1fb="RPCT-23"  
1..12805  
/note="assembly\_fragment:1158"  
12906..67118  
/note="assembly\_fragment:201"  
67219..99173  
/note="assembly\_fragment:2057"  
99274..111725  
/note="assembly\_fragment:2060"  
111826..115810  
/note="assembly\_fragment:225"  
115911..139424  
/note="assembly\_fragment:2668"  
139525..156165  
/note="assembly\_fragment:278"  
156266..170601  
/note="assembly\_fragment:4020"  
170702..177014  
/note="assembly\_fragment:4351"  
BASE COUNT 52674 a 37604 c 36695 g 49232 t 809 others  
ORIGIN

Query Match 1.1%; Score 21; DB 2; Length 177014;  
Best Local Similarity 100.0%; Pred. No. 4.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 299 gaagactatctcaagcaagc 319  
 |||  
 Db 116245 GAAGACTATCTCAAGCAAGC 116265

RESULT 14  
 AL611984/c 206591 bp DNA linear HTG 13-OCT-2001  
 LOCUS Mus musculus chromosome 4 clone RP23-41608, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, in unordered pieces.  
 ACCESSION AL611984  
 VERSION AL611984.6 GI:16151499  
 KEYWORDS HTG: HTGS.PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 206591)  
 REFERENCE Direct Submission  
 AUTHORS Submitted (12-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 TITLE CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 JOURNAL requests: clonerequest@sanger.ac.uk  
 On Oct 15, 2001 this sequence version replaced gi:16116600.  
 COMMENT ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 Project Information  
 Center project name: BM416J8  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 203072 bases at least Q40  
 Consensus quality: 203785 bases at least Q30  
 Consensus quality: 204261 bases at least Q20  
 Insert size: 204991; sum-of-contigs  
 Insert size: 196126; 3.5% error; agarose-fp  
 Quality coverage: 9.25x in Q20 bases; sum-of-contigs Quality  
 coverage: 9.86x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

## FEATURES

source 1..206591  
 location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="4"  
 /clone="RP23-416J8"  
 /clone\_1lb="RPC1-23"  
 1..12340  
 /note="assembly\_fragment:01861  
 fragment\_chain:1"  
 misc\_feature 1241..14533  
 /note="assembly\_fragment:00110  
 fragment\_chain:1"  
 misc\_feature 14634..25420  
 /note="assembly\_fragment:02706  
 fragment\_chain:1"  
 misc\_feature 25521..38180  
 /note="assembly\_fragment:02641  
 fragment\_chain:1"  
 misc\_feature 38281..45684  
 /note="assembly\_fragment:01195  
 fragment\_chain:1"  
 misc\_feature 45785..49116  
 /note="assembly\_fragment:00318

fragment\_chain:1"  
 49217..103493  
 /note="assembly\_fragment:01418  
 fragment\_chain:1"  
 misc\_feature 103594..122215  
 /note="assembly\_fragment:04869  
 fragment\_chain:1"  
 misc\_feature 122316..125353  
 /note="assembly\_fragment:00804  
 fragment\_chain:1"  
 misc\_feature 125454..127674  
 /note="assembly\_fragment:02278  
 fragment\_chain:1"  
 misc\_feature 127775..132117  
 /note="assembly\_fragment:01785  
 fragment\_chain:2"  
 misc\_feature 132218..139371  
 /note="assembly\_fragment:03689  
 fragment\_chain:2"  
 misc\_feature 139472..155577  
 /note="assembly\_fragment:02580  
 fragment\_chain:3"  
 misc\_feature 15678..188900  
 /note="assembly\_fragment:04111  
 fragment\_chain:3"  
 misc\_feature 189001..195856  
 /note="assembly\_fragment:00631  
 fragment\_chain:4"  
 misc\_feature 195957..198542  
 /note="assembly\_fragment:03866  
 fragment\_chain:4"  
 misc\_feature 198643..206591  
 /note="assembly\_fragment:02750  
 fragment\_chain:4  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 53779 a 47056 c 47603 g 56551 t 1602 others  
 ORIGIN

Query Match 1.1%; Score 21; DB 2; Length 206591;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1044 cccgcgcgatatctctcaaaaa 1064  
 |||  
 Db 156967 CCCTGCCGATATCTCTAAAAA 156947

## RESULT 15

AL645546 225609 bp DNA linear HTG 13-NOV-2001  
 LOCUS Mus musculus chromosome 13 clone RP23-124G18, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, in unordered pieces.  
 ACCESSION AL645546  
 VERSION AL645546.4 GI:16944328  
 KEYWORDS HTG: HTGS.PHASE1.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (sites)  
 REFERENCE Direct Submission  
 AUTHORS Submitted (08-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,  
 TITLE Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 JOURNAL humquerry@sanger.ac.uk  
 On Nov 15, 2001 this sequence version replaced gi:16754763.  
 COMMENT ----- Genome Center  
 Center: UK Medical Research Council  
 Center code: UK-MRC  
 Web site: http://mrcseq.har.mrc.ac.uk  
 Contact: mouse@har.mrc.ac.uk

```

----- Project Information
Center project name: bm124G18
----- Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 224286 bases at least Q40
Consensus quality: 22463 bases at least Q30
Insert size: 225409; sum-of-contigs
Insert size: 216124; 7.8% error; agarose-fp
Quality coverage: 11.91x in Q20 bases; sum-of-contigs Quality
coverage: 13.08x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence..
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
    Location/Qualifiers
        1..225609
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /chromosome="13"
        /clone="RP23-124G18"
        /clone_1fb="RPC1-23"
        1..101657
        /note="assembly_fragment:01246"
        clone_end:T7
        vector_side:left"
        101758..190829
        /note="assembly_fragment:02144.0"
        190930..225609
        /note="assembly_fragment:03429"
        clone_end:SP6
        vector_side:right"
BASE COUNT 66198 a 45409 c 44739 g 69047 t 216 others
ORIGIN
Query Match 1.1%; Score 21; DB 2; Length 225609;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 acacagcagccttagcacaca 238
|||||
Db 147727 ACACAGCAGCCTTAGCACACA 147747

```

Search completed: August 20, 2002, 03:20:54  
 Job time: 7604 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 04:12:06 ; Search time 28.88 Seconds

(without alignments)  
2184.554 Million cell updates/sec

Title: US-09-662-812-2

Perfect score: 568  
Sequence: 1 MGLFHLFLGLLCSLPISL.....TEGDAKNPVLGLIKIIVE 568

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	100.0	568	22	AAV72972
2	327	57.6	348	20	AAV34620
3	241	42.4	245	20	AAV34619
4	12	2.1	468	20	AAV36830
5	10	1.8	435	22	AAU35103
6	8	1.4	52	22	AAV66330
7	8	1.4	77	22	AAV99877
8	8	1.4	94	22	AAV80047
9	8	1.4	104	22	AAV29407
10	8	1.4	109	22	AAV37519
11	8	1.4	142	21	AAV41090

12	8	1.4	207	22	ABG06236	Novel human diago
13	8	1.4	254	22	ABG14349	Novel human diago
14	8	1.4	372	20	AAW89449	A gida2 polypeptid
15	8	1.4	379	21	AAV10364	Arabidopsis thalia
16	8	1.4	382	21	AAV10363	Arabidopsis thalia
17	8	1.4	401	22	AAV38082	Streptococcus pneu
18	8	1.4	411	16	AAV67733	Artematic dihydridol
19	8	1.4	433	22	AAV34397	Staphylococcus aur
20	8	1.4	435	20	AAW89448	A gida2 polypeptid
21	8	1.4	435	22	AAV37266	Staphylococcus aur
22	8	1.4	444	20	AAW89450	A gida2 polypeptid
23	8	1.4	444	22	AAV37761	Streptococcus pneu
24	8	1.4	447	22	AAV98357	Escherichia coli p
25	8	1.4	448	22	AAV90237	C glutamicum prote
26	8	1.4	448	22	AAW80046	Corynebacterium g1
27	8	1.4	498	22	AAV03145	Streptococcus pyog
28	8	1.4	520	22	AAV41718	Propionibacterium
29	8	1.4	523	21	AAV09513	Arabidopsis thalia
30	8	1.4	523	21	AAV38975	Arabidopsis thalia
31	8	1.4	528	21	AAV09512	Arabidopsis thalia
32	8	1.4	558	21	AAV38974	Arabidopsis thalia
33	8	1.4	568	21	AAV15185	Partial soybean ze
34	8	1.4	568	21	AAV15186	Wheat zeta-caroten
35	8	1.4	640	22	AAV03549	Novel human diago
36	7	1.2	57	22	AAV2739	Thrombin immunogen
37	7	1.2	34	21	AAV28021	Human secreted pro
38	7	1.2	52	22	AAV44489	Peptide #11995 enc
39	7	1.2	58	22	AAV65610	Human brain expres
40	7	1.2	58	22	AAV78237	Human bone marrow
41	7	1.2	58	22	AAV8237	Peptide #12537 enc
42	7	1.2	58	22	AAV38500	Human prostate can
43	7	1.2	61	21	AAV56502	Leishmania antigen
44	7	1.2	69	19	AAV70258	Drosophila melanog
45	7	1.2	73	22	AAV68835	

#### ALIGNMENTS

RESULT 1  
AAV72972 standard; Protein: 568 AA.  
AC AAV72972;  
XX  
XX 13-JUN-2001 (first entry)  
DE Chlamydia pneumoniae outer membrane protein (OMP).  
XX outer membrane protein; OMP; antibacterial; vaccine; gene therapy;  
KW antibody; medicament; Chlamydia infection.  
XX  
OS Chlamydia pneumoniae.  
XX  
XX WO200121804-A1.  
XX  
XX 29-MAR-2001.  
PD  
PF 15-SEP-2000; 2000WO-CA01088.  
XX  
XX 20-SEP-1999; 99US-0154652.  
XX  
XX (AVET ) AVENTIS PASTEUR LTD.  
PA  
XX Murdin AD, Oomen RP, Wang J, Dunn P;  
PI  
XX WPI; 2001-244939/25.  
DR  
XX N-PSDB; AAD03024.  
XX  
XX Novel Chlamydia pneumoniae outer membrane protein and polynucleotides  
PT encoding them, useful as components of vaccines for treating Chlamydia  
PT infections, and for detecting Chlamydia infections in the body fluids  
PT of mammals -

XX Claim 16; Fig 1; 82pp; English.  
XX  
XX The present sequence is Chlamydia pneumoniae OMP (outer membrane  
CC protein). OMP is useful in the production of vaccines, antibodies  
CC and pharmaceutical compositions which are useful for treating or  
CC preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or  
CC C. pecorum) infections. They are also useful as diagnostic reagents for  
CC detecting Chlamydia infection which involves assaying a body fluid  
CC of a mammal to be tested for the components. The OMP vaccine is  
CC useful in the preparation of a medicament for preventing and/or  
CC treating Chlamydia infection. The primers derived from OMP gene are  
CC also useful for detecting and/or identifying Chlamydia in a biological  
CC material. OMP antibodies are also useful as reagents for purifying OMP  
CC from a biological sample which involves carrying out antibody-based  
CC affinity chromatography with the biological sample. OMP gene is  
CC also useful in gene therapy.  
XX  
XX Sequence 568 AA:  
SO  
Query Match 100.0%; Score 568; DB 22; Length 568;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGLFHLTFGLGSLSPISLVAKPESVGHKILYSTOSTQOALATYLEALDAYGDHDF 60  
DB 1 mglfhlftlgllscslpislvakfpesvghkillyststqgalatylealdygdhdf 60  
QY 61 VLKRIGEDYLKOSIHSSPQTRKSTIIIGAGLAGSSEALDVLSQAMETADPILOQLVLISAV 120  
DB 61 vlkrigedylkqshsspqtrkstiiigaglagsssealdvlsqametadpiqlvlisav 120  
QY 121 SGHLGKTSDDLFFALASPVYIRLEAAYRLANKNTKVYIDHLSFIKRLPEEIOCLSA 180  
DB 121 sghlgktsddllffalaspvyirleaaayrlankntkvihlsfihkrlpeelqclsa 180  
QY 121 sghlgktsddllffkalaapvyirleaaayrlankntkvihlsfihkrlpeelqclsa 180  
DB 121 sghlgktsddllffkalaapvyirleaaayrlankntkvihlsfihkrlpeelqclsa 180  
QY 181 IFRLRTEESDAYTRDLAAKKSAIRSATAOIGFYOOKRFLPTIRNLNTSASPDQDAI 240  
DB 181 iflrletesdaytrdllaakksairstataiqeygqkrlfpltrnlltsaspdqda 240  
QY 241 LVALGKLDGOSYNIKKOLKPDVYTLAAAOALIALGKEEDALPVKKOALEERPAAL 300  
DB 241 lvalgkldgqsyynikkqlkpdvctlaaaqalialgkeedaipvkkqaleerpal 300  
QY 301 YALRHLPEISIGIPALPIFLKTKNSEAKLNVALLLELGCPTPKLLEYTERLYOPHYNE 360  
DB 301 yalrhlpseigipalpiflktnseaklnvalalalelgcptpkllleyterlyvphne 360  
QY 361 TLALSFSGKRLQWKKRNITIVPODPRERLSTTRGLEQILFTFLRPREKAVLPCTY 420  
DB 361 tlaistfskgrtlqwmkrrnitiivpodprerlstrrgleeqilftflrpkvaylpcty 420  
QY 421 KLILSOKROLTTATISFISHSHOEALDLFOAKLGPSEPIIRAYADAIYNTLTPDEPKK 480  
DB 421 klilsokrolttatishshshoealldlfoaklgpsepiirayadaiyntltpdepkk 480  
QY 481 RSLHDYAKKLIOETLLEFVDTENORPHPSMPYRLRYOVTPESRKLMLDLLETATSKSEED 540  
DB 481 rslhdyakklloetllefvdtenorhpsmpyrlryovtpesrklmldlletatskssed 540  
QY 541 IRLLIQMTGEDANKFPVLAGILIKIIVE 568  
DB 541 irllliqmtgedankfpvlagililike 568  
RESULT 2  
ID AAY34620 standard; Protein: 348 AA.  
XX AAY34620;  
XX AC  
XX 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae lipoprotein sequence.  
DE  
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX Chlamydia pneumoniae.  
XX  
XX MO9927105-A2.  
XX  
XX 03-JUN-1999.  
XX  
XX 20-NOV-1998; 98WO-IB01890.  
XX  
XX 04-NOV-1998; 98US-0107078.  
XX  
XX 21-NOV-1997; 97FR-0014673.  
XX  
XX (GEST ) GENSET.  
XX  
XX Griffiths R;  
XX  
XX WPI: 1999-357842/30.  
XX  
XX Genome sequence of Chlamydia pneumoniae  
XX  
XX Page 640; Disclosure: 1912pp; English.  
XX  
XX AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
XX Sequence 348 AA:  
SO  
Query Match 57.6%; Score 327; DB 20; Length 348;  
Best Local Similarity 100.0%; Pred. No. 6.7e-302;  
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGLFHLTFGLGSLSPISLVAKPESVGHKILYSTOSTQOALATYLEALDAYGDHDF 60  
DB 21 mglfhlftlgllscslpislvakfpesvghkillyststqgalatylealdygdhdf 80  
QY 61 VLKRIGEDYLKOSIHSSPQTRKSTIIIGAGLAGSSEALDVLSQAMETADPILOQLVLISAV 120  
DB 61 vlkrigedylkqshsspqtrkstiiigaglagsssealdvlsqametadpiqlvlisav 140  
QY 121 SGHLGKTSDDLFFALASPVYIRLEAAYRLANKNTKVYIDHLSFIKRLPEEIOCLSA 180  
DB 121 sghlgktsddllffalaspvyirleaaayrlankntkvihlsfihkrlpeelqclsa 200  
QY 181 IFRLRTEESDAYTRDLAAKKSAIRSATAOIGFYOOKRFLPTIRNLNTSASPDQDAI 240  
DB 181 iflrletesdaytrdllaakksairstataiqeygqkrlfpltrnlltsaspdqda 260  
QY 241 LVALGKLDGOSYNIKKOLKPDVYTLAAAOALIALGKEEDALPVKKOALEERPAAL 300  
DB 241 lvalgkldgqsyynikkqlkpdvctlaaaqalialgkeedaipvkkqaleerpal 320  
QY 301 YALRHLPEISIGIPALPIFLKTKNSEA 327  
DB 321 yalrhlpseigipalpiflktnsea 347  
RESULT 3



AAV34619  
 ID AAV34619 standard; Protein; 245 AA.  
 XX  
 AC AAV34619;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Chlamydia pneumoniae transmembrane protein sequence.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN MO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffais R;  
 XX  
 DR WPI; 1999-357842/30.  
 XX  
 PS Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Page 639-640; Disclosure; 1912pp; English.  
 XX  
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 CC  
 XX  
 SQ Sequence 245 AA;  
 XX

Query Match 42.4%; Score 241; DB 20; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-220;  
 Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 328 KLVNALALDELGCDTPKLEYTERLVQPHYNETLALFSKGRTLQWKRNVIIIVPDDPQ 387  
 DB 5 klvnaiallalgcdtpkleyterlvqphynetlalfskgrtlqwmkrnviiivpddpq 64

OY 388 EREBLSTGTGLEQLITFLPRPKKAYLPCIVYLLASQKQTALTAISPLSHSHQBAL 447  
 DB 65 ereblstgtgleqlitflprpkkaylpcivyllasqkqtaltaisfshshqbal 124

OY 448 DLLEFOAAKLGEPIIRAYADLAIVNLTKDEPKRSLHDYAKKLIOETLLFVDTENQRPH 507  
 DB 125 dllefoaaklgepiirayadlaivnltkdepkrsldyakkliioetllfvdtengrph 184

OY 508 SMPYLARQVPESEKTKMLDILETLATSKSSEDIRLLIQMTGEDARNFPVLAQLIKIV 567  
 DB 185 smpylarqvpesektkmldiletlatsksseidirlliqmtgedarnfpvlaglikiv 244

OY 568 E 568  
 DB 245 e 245

RESULT 4  
 AAY36830  
 ID AAY36830 standard; Protein; 468 AA.  
 XX  
 AC AAY36830;  
 XX  
 DT 07-OCT-1999 (first entry)  
 XX  
 DE Chlamydia trachomatis lipoprotein sequence.  
 XX  
 KW Vaccine; eye disease; conventional trachoma; nongonococcal trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perinephalitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9928475-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98WO-IB01939.  
 XX  
 PR 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffais R;  
 XX  
 DR WPI; 1999-371125/31.  
 XX  
 PS Genome sequence of Chlamydia trachomatis  
 XX  
 PS Disclosure; Page 708-710; 1755pp; English.  
 XX  
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AAY201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nongonococcal trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC perinephalitis, bartholinitis; pneumopathy in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 CC  
 XX  
 SQ Sequence 468 AA;  
 XX

Query Match 2.1%; Score 12; DB 20; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 0.012;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 141 PVIRLEAAYRLA 152  
 DB 142 pvirleaayrlla 153

RESULT 5  
 AAU35103  
 ID AAU35103 standard; Protein; 435 AA.  
 XX  
 AC AAU35103;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Enterococcus faecalis cellular proliferation protein #390.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.  
 XX Enterococcus faecalis.  
 OS  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PE 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI: 2001-611495/70.  
 DR N-PSDB; AAS52962.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 10696; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 435 AA;  
 XX

Query Match 1.8%; Score 10; DB 22; Length 435;  
 Best Local Similarity 100.0%; Pred. No. 0.93;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 85 TTIGAGTAGS 94  
 Db 6 tligagtag 15  
 XX

RESULT 6  
 ID AAB66330 standard; Protein: 52 AA.  
 XX  
 AC AAB66330;  
 XX  
 DT 05-APR-2001 (first entry)  
 XX  
 DE Synecocystis zeta-carotene desaturase SEQ ID NO: 78.  
 XX

KW Linoleate isomerase; conjugated linoleic acid; conjugated linolenic acid;  
 KW CIA: oil; lipase-like protein; acetyltransferase-like protein.  
 XX  
 OS Synecocystis sp.  
 XX  
 PN WO200100846-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PE 30-JUN-2000; 2000WO-US18149.  
 XX  
 PR 30-JUN-1999; 99US-0141798.  
 PR 28-APR-2000; 2000US-0561077.  
 XX  
 PA (DCVB-) DCV INC DBA BIO-TECH RESOURCES.  
 XX  
 PI Rosson RA, Deng M, Grund AD, Peng SS;  
 XX  
 DR WPI: 2001-102894/11.  
 XX  
 PT Isolated linoleate isomerase proteins and nucleic acids, useful in  
 PT methods for the production of conjugated linoleic acid or conjugated  
 PT linolenic acid (CIA) -  
 XX  
 PS Example 13; Page 242; 244pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences for the  
 CC linoleate isomerase enzyme from Propionibacterium acnes. This enzyme is  
 CC useful for producing conjugated linoleic and linolenic acids (CIAs) from  
 CC linoleic and linolenic acids in oils. Also provided are the sequences of  
 CC an acetyltransferase-like protein and a lipid-like protein from the same  
 CC organism.  
 XX  
 SQ Sequence 52 AA;  
 XX

Query Match 1.4%; Score 8; DB 22; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 86 TTIGAGTAG 93  
 Db 4 tligagtag 11  
 XX

RESULT 7  
 ID AAB99877 standard; Protein: 77 AA.  
 XX  
 AC AAB99877;  
 XX  
 DT 21-SEP-2001 (first entry)  
 XX  
 DE Physcomitrella patens 42\_ck10\_g09fwd protein.  
 XX  
 KW Tocopherol and carotenoid metabolism related protein; TCMRP; synchysis;  
 KW Physcomitrella patens; moss; algae; microorganism; fungus; plant;  
 KW identification; genome mapping; modulation; evolutionary study;  
 KW cellular production; fine chemical.  
 XX  
 PA Physcomitrella patens.  
 XX  
 PN WO200144276-A2.  
 XX  
 PD 21-JUN-2001.  
 XX  
 PE 14-DEC-2000; 2000WO-EP12698.  
 XX  
 PR 16-DEC-1999; 99US-0171121.  
 XX  
 PA (BADI ) BASF PLANT SCI GMBH.  
 XX  
 DE Lerchl J, Renz A, Ehrhardt T, Reindl A, Cirpus P, Bischoff F;  
 XX

PI Frank M, Freund A, Duwenig E, Schmidt R, Reeki R, Badur R;  
XX WPI: 2001-398121/42.  
DR N-PSDB; AAH44250.  
XX  
PT Tocopherol and carotenoid metabolism related protein (TCMRP), used to  
PT produce fine chemicals, is isolated from mosses, algae, microorganisms,  
PT fungi, plants, or their fragments -  
XX  
PS Claim 28; Page 119; 123pp; English.  
XX  
CC The present invention describes isolated tocopherol and carotenoid  
CC metabolism related proteins (TCMRP) (I) from mosses or algae,  
CC microorganisms or fungi, plants, or its fragments. (I) can be used as  
CC enzymes in the production of fine chemicals or in the metabolism of  
CC tocopherol and carotenoids. (I) also assist in transmembrane transport.  
CC The fine chemicals that can be produced include lipids, fatty acids,  
CC vitamins, cofactors, enzymes, amino acids, and nucleotide bases.  
CC Nucleotide sequences, proteins, vectors and host cells from the present  
CC invention can be used: (a) to identify mosses related to Physcomitrella  
CC patens; (b) in mapping genomes of mosses related to Physcomitrella  
CC studies; (c) in the modulation of TCMRP activity; (d) in evolutionary  
CC studies; (e) in the determination of functional TCMRP regions; (f) and  
CC in the cellular production of fine chemicals. AAH44222 to AAH44262  
CC encode the Physcomitrella patens TCMRP proteins given in AAB9849 to  
CC AAB9889. AAH44212 to AAH44221 represent nucleotide sequence used in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 77 AA;  
  
Query Match 1.4%; Score 8; DB 22; Length 77;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 86 ITGAGLAG 93  
DB 49 IIGAGLAG 56  
  
RESULT 8  
AAB80047  
ID AAB80047 standard; Protein; 94 AA.  
XX  
AC AAB80047;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:828.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polypeptide; enzyme.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200100843-A2.  
PD  
PD 04-JAN-2001.  
XX  
XX  
PF 23-JUN-2000; 2000MO-IB00923.  
XX  
XX  
XX 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.  
PR 31-AUG-1999; 99DE-1041394.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
XX (BADI ) BASF AG.  
XX  
XX Pompejus M, Kroegeer B, Schroeder H, Zeider O, Haberhauer G;  
PI WPI: 2001-137957/14.  
DR N-PSDB; AAF72166.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX  
PS Claim 20; Page 1301-1302; 1737pp; English.  
XX  
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC MP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polypeptides and enzymes.  
XX  
SQ Sequence 94 AA;

Query Match 1.4%; Score 8; DB 22; Length 94;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 86 IIGAGLAG 93  
|||||||  
Db 5 IIGAGLAG 12

RESULT 9  
ABG29407  
ID ABG29407 standard; Protein: 104 AA.  
XX AC ABG29407;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #29398.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB: AAS93594.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -  
XX PS Claim 20; SEQ ID No 59766; 103pp; English.  
XX PS The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 104 AA;

Query Match 1.4%; Score 8; DB 22; Length 104;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 88 GAGLAGSS 95  
|||||||  
Db 83 GAGLAGSS 90

RESULT 10  
AAU37519  
ID AAU37519 standard; Protein: 109 AA.  
XX AC AAU37519;  
XX DT 14-FEB-2002 (first entry)  
XX DE Staphylococcus aureus cellular proliferation protein #1689.  
XX KW Antisense; prokaryotic cellular proliferation protein;  
XX KW antibiotic; antibacterial; drug design.  
XX OS Staphylococcus aureus.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207727P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX PI Yamamoto RT, Xu HH;  
XX DR WPI; 2001-611495/70.  
XX DR N-PSDB: AAS55378.  
XX PT New polynucleotides for the identification and development of  
XX PT antibiotics, comprise sequences of antisense nucleic acids -  
XX PS Example 3; Seq ID No 13112; 51pp; English.  
XX PS The invention relates to antisense inhibitors of genes essential to  
XX CC prokaryotic cellular proliferation, their use in identifying the  
XX CC genes, their use in the discovery of novel antibiotics, the essential  
XX CC genes themselves and the encoded proteins. The prokaryotes used are  
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX CC invention is also useful for the identification of potential new targets  
XX CC for antibiotic development. The antisense nucleic acids can also be used  
XX CC to identify proteins used in proliferation, to express these proteins,  
XX CC and to obtain antibodies capable of binding to the expressed proteins.  
XX CC The proteins can be used to screen compounds in rational drug discovery  
XX CC programmes. The antisense nucleic acid sequence is also useful to screen  
XX CC for homologous nucleic acids which are required for cell proliferation in  
XX CC a wide variety of organisms. The present sequence represents an  
XX CC essential prokaryotic cellular proliferation protein.  
XX CC Note: The sequence data for this patent did not form part  
XX CC of the printed specification, but was obtained in electronic  
XX CC format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 109 AA;

Query Match 1.4%; Score 8; DB 22; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 IGAGLAGS 94  
 |||||  
 Db 8 igaglags 15

RESULT 11  
 AAB41090  
 ID AAB41090 standard; Protein; 142 AA.  
 XX  
 AC AAB41090;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF854 polypeptide sequence SEQ ID NO:1708.  
 XX  
 KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnerable; antiparasitic; antiparkinsonian; noctropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;  
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PE 31-MAR-2000; 2000MO-US08621.  
 XX  
 PF 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Leach M;  
 DR WPI; 2000-602362/57.  
 DR N-PSDB; AAC75299.  
 XX  
 PT Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 11; Page 1359; 5507pp; English.

AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 sequences have activities such as: cytostatic; hepatotropic; vulnerable;  
 antiparasitic; antiparkinsonian; noctropic; neuroprotective;  
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
 CC antihypoid; and antianaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The

CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation, to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 142 AA:

Query Match 1.4%; Score 8; DB 21; Length 142;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 IIGAGLAG 93  
 |||||  
 Db 28 iigaglag 35

RESULT 12  
 ABG06236  
 ID ABG06236 standard; Protein; 207 AA.  
 XX  
 AC ABG06236;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #6227.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PE 30-MAR-2001; 2001MO-US08631.  
 XX  
 PF 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS70423.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID NO 36595; 103pp; English.

The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
SQ Sequence 207 AA;

Query Match 1.4%; Score 8; DB 22; Length 207;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 GAGLAGSS 95  
|||||  
Db 86 gaglagss 93

RESULT 13  
ABG14349  
ID ABG14349 standard; Protein: 254 AA.  
XX  
AC ABG14349;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #14340.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
PI  
DR WPI: 2001-639362/73.  
DR N-PSDB; AAS78536.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
PT  
PS Claim 20; SEQ ID No 44708; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
SQ Sequence 254 AA;

Query Match 1.4%; Score 8; DB 22; Length 254;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 88 GAGLAGSS 95  
|||||  
Db 56 gaglagss 63

RESULT 14  
AAM89449  
ID AAM89449 standard; Protein: 372 AA.  
XX  
AC AAM89449;  
XX  
DT 18-MAR-1999 (first entry)  
XX  
DE A gida2 polypeptide fragment.  
XX  
KW gida2; Staphylococcus aureus WCUH29; bacterial infection;  
KW Helicobacter pylori infection; cancer; ulcer; gastritis; antibacterial;  
KW wound treatment; bacterial adhesion; matrix protein.  
XX  
OS Staphylococcus aureus.  
XX  
PN EP889131-A2.  
XX  
PD 07-JAN-1999.  
XX  
PF 30-JUN-1998; 98EP-0305203.  
XX  
PR 01-JUL-1997; 97US-0051380.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PA (SMIK ) SMITHKLINE BEECHAM PLC.  
PI Burnham M, Debouck CM, Kallender H, Lenox AL, Mooney JL;  
PI Palmer LM, Zhong Y;  
PI  
DR WPI: 1999-062662/06.  
DR N-PSDB; AAV82085.  
XX  
PT New isolated gida2 polypeptide from Staphylococcus aureus - used to  
PT diagnose, treat and prevent bacterial infections e.g. S. aureus and  
PT H. pylori, related cancers, ulcers and gastritis and to prevent  
PT adhesion of bacteria to matrix proteins  
PT  
PS Claim 1; Page 6; 41pp; English.  
XX  
XX The present sequence represents a gida2 protein fragment of  
CC Staphylococcus aureus WCUH29. The gida2 proteins and nucleic acids  
CC are used to treat conditions requiring increased activity or expression  
CC of gida2, while conditions (e.g. bacterial infections) requiring  
CC inhibition of such activity or expression are treated by administering  
CC an antagonist, inhibitory nucleic acid or competitive polypeptide.  
CC Infection by S. aureus is treated, but also Helicobacter pylori  
CC infections and related cancers, ulcers and gastritis. These antibacterial  
CC agents may also be used to treat in-dwelling devices to prevent infection  
CC or generally as wound treatments to prevent adhesion of bacteria to  
CC matrix proteins.

SQL Sequence 372 AA;

Query Match 1.4%; Score 8; DB 20; Length 372;  
Best Local Similarity 100.0%; Pred No. 64;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 IGAGLAGS 94  
|||||||  
Db 8 igaglags 15

RESULT 15  
AAGI0364  
ID AAGI0364 standard; Protein; 379 AA.  
XX  
AC AAGI0364;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8654.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0133284.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 19-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145519.  
PR 28-JUL-1999; 99US-0145521.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 1.4%; Score 8; DB 21; Length 379;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 IIGAGLAG 93  
db 51 IIGAGLAG 58

Search completed: August 20, 2002, 04:25:18  
Job time: 792 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: August 20, 2002, 03:12:10 ; Search time 58.23 Seconds  
(without alignments)  
1083.461 Million cell updates/sec

Title: US-09-662-812-2  
Perfect score: 2833  
Sequence: 1 MGFLHFLFLGLLCSLPISL.....TEGDKNFVYLAGLLIKIVE 568

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
A.Geneseq\_032802:\*

1:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/genesqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2833	100.0	568	22	AAV72972	Chlamydia pneumoni
2	1619	57.1	348	20	AAV34620	Chlamydia pneumoni
3	1215	42.9	245	20	AAV34619	Chlamydia pneumoni
4	1092	38.5	468	20	AAV36830	Chlamydia trachoma
5	296	10.4	96	20	AAV37834	Amino acid sequenc
6	137.5	4.9	1015	22	ABG04636	Novel human diagno
7	132.5	4.7	514	22	ABR64398	Drosophila melanog
8	129.5	4.6	7201	22	ABR71136	Drosophila melanog
9	128.5	4.5	1870	22	ABG21017	Novel human diagno
10	127.5	4.5	2017	22	ABG06301	Novel human diagno
11	126	4.4	631	17	AAV9675	RHAMM 1-2a isoform

12	123.5	4.4	606	17	AAV9673	Receptor for hyalu
13	123.5	4.4	630	18	AAW39166	Mouse RHAMM protei
14	121.5	4.3	1838	22	AAU00015	Human plexin-B2.
15	121.5	4.3	1849	22	ABH11818	Human plexin-B1/SE
16	121.5	4.3	2985	19	AAW56643	DNA-dependent prot
17	121.5	4.3	3960	19	AAW56642	Protein Kinase cat
18	119.5	4.2	688	22	AAV64567	Human vacuole prot
19	118.5	4.2	716	22	AAV93421	Human polypeptide,
20	118.5	4.2	2117	22	ABW59379	Drosophila melanog
21	118.5	4.2	2117	22	ABW67222	Drosophila melanog
22	117.5	4.1	1111	22	ABG09489	Novel human diagno
23	117.5	4.1	1336	22	AAV60489	Human cell cycle a
24	117.5	4.1	2688	22	AAV40883	Human polypeptide
25	116.5	4.1	740	22	ABW60942	Drosophila melanog
26	115.5	4.1	1111	22	AAV80108	Human protein SEQ
27	115.5	4.1	1111	22	AAV80109	Human protein SEQ
28	115.5	4.1	1569	22	ABG15232	Novel human diagno
29	115.5	4.1	1743	19	AAV98879	H. pylori GHPD 175
30	115	4.1	1185	19	AAV98508	H. pylori GHPD 123
31	114.5	4.0	883	22	ABW59757	Drosophila melanog
32	114	4.0	746	21	AAV42530	Arabidopsis thalia
33	114	4.0	840	21	AAV42529	Arabidopsis thalia
34	114	4.0	932	21	AAV42528	Arabidopsis thalia
35	114	4.0	1179	22	AAV01107	CBE 110 protein se
36	114	4.0	1217	22	AAV52358	Putative TBP inter
37	114	4.0	1388	19	AAV56475	Protein with Rho p
38	114	4.0	2076	22	AAV34319	Staphylococcus aur
39	114	4.0	2186	22	AAV37320	Staphylococcus aur
40	114	4.0	2633	22	AAV39097	Human polypeptide
41	113.5	4.0	1119	22	AAV79124	Human protein SEQ
42	113.5	4.0	2207	22	AAV32041	Novel human secret
43	113	4.0	1179	22	AAV37669	Streptococcus pneu
44	112.5	4.0	1690	22	ABW61144	Drosophila melanog
45	112.5	4.0	1690	22	ABW61173	Drosophila melanog

## ALIGNMENTS

RESULT 1  
AAV72972 standard; Protein; 568 AA.  
ID AAV72972 standard; Protein; 568 AA.  
XX  
AC AAV72972;  
XX  
DT 13-JUN-2001 (first entry)  
XX  
DE Chlamydia pneumoniae outer membrane protein (OMP).  
XX  
KW Outer membrane protein; OMP; antibacterial; vaccine; gene therapy;  
KW antibody; medicament; Chlamydia infection.  
XX  
OS Chlamydia pneumoniae.  
XX  
PN WO200121804-A1.  
XX  
PD 29-MAR-2001.  
XX  
PF 15-SEP-2000; 2000MO-CA01088.  
XX  
PR 20-SEP-1999; 99US-0154652.  
XX  
PA (AVER ) AVENTIS PASTEUR LTD.  
XX  
PI Murdin AD, Oomen RP, Wang J, Dunn P;  
XX  
DR WPI: 2001-244939/25.  
XX  
N-PSDB: AAD03024.  
XX  
PT Novel Chlamydia pneumoniae outer membrane protein and polynucleotides  
PT encoding them, useful as components of vaccines for treating Chlamydia  
PT infections, and for detecting Chlamydia infections in the body fluids  
of mammals -

XX Claim 16; Fig 1; 82pp; English.

CC The present sequence is Chlamydia pneumoniae OMP (outer membrane  
CC protein). OMP is useful in the production of vaccines, antibodies  
CC and pharmaceutical compositions which are useful for treating or  
CC preventing Chlamydia (C. trachomatis, C. psittaci, C. pneumoniae or  
CC C. pecorum) infections. They are also useful as diagnostic reagents for  
CC detecting Chlamydia infection which involves assaying a body fluid  
CC of a mammal to be tested for the components. The OMP vaccine is  
CC useful in the preparation of a medicament for preventing and/or  
CC treating Chlamydia infection. The primers derived from OMP gene are  
CC also useful for detecting and/or identifying Chlamydia in a biological  
CC material. OMP antibodies are also useful as reagents for purifying OMP  
CC from a biological sample which involves carrying out antibody-based  
CC affinity chromatography with the biological sample. OMP gene is  
CC also useful in gene therapy.

CC Sequence 568 AA;

Query Match 100.0%; Score 2833; DB 22; Length 568;

Best Local Similarity 100.0%; Pred. No. 1,1e-227; Mismatches 0; Indels 0; Gaps 0;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MGLFHLTLFGLLCSLPISLVAKPEESVGHKILYISTOSTOQALATYLEALDAYGDHDF 60
DB 1 mglfhltlflllcslpislvakfpesvghkilyistgstqalatyaleadaygdhdf 60
OY 61 VLKRGEDYLNKSHSSDPQTRKSTITIGAGLSSEALDVLVSQAMETADPLOQLVLSAV 120
DB 61 vlkrigedylnkshssdpqtrkstltaglsgsealdvlsqametaadpqlqlvlsav 120
OY 121 SGHGLKTSDDLFFKALASPYVIRLEAAYRLANKTKFVIDHLSFTHKLPREIQCISAA 180
DB 121 sghglktsddlffkalsapyvirleaaayrlanlknkvidhlsfthklpreeicisaa 180
OY 181 IFLRLETESDAYIRDLAARSAIRSATALQIGEYQOKRFLPRLNLTSSASPODOEAI 240
DB 181 iflrleteesdayirldlaarsaaisrsatalqigeyqokrflprlnltsaspdodeai 240
OY 241 IFLRLETESDAYIRDLAARSAIRSATALQIGEYQOKRFLPRLNLTSSASPODOEAI 240
DB 241 iflrleteesdayirldlaarsaaisrsatalqigeyqokrflprlnltsaspdodeai 240
OY 241 LVALGKLLKDGSGSYNNIKKQLOKPDVDTLAAQAALIALGKEEDALPVYIKKQALEERPRAL 300
DB 241 lvalgkllkdsgsyyinnikkqlokpdydtltaaqaalialgkeedalpvyikkqaleerpral 300
OY 301 YALRHLPESEIGIPALPIFLTKNSEAKLVNALLELGCDFPRKLETERIVQPHNE 360
DB 301 yalrhlpseigipalpifltknseaklvnalallelgcdfprkleyterlvqphne 360
OY 361 TLALSFSGKRTLQWKKRVNITVPDPOERERLSTTRGLEBOILFTFLRLEKAVLPCTY 420
DB 361 tlaifsfsgkrtlwkkrvnitvpdpoeerlstrgleboilftflrlepkeavlpcty 420
OY 421 KLASQRTQALATTAISPLSHTSHOBALDLFOAAKLECEPIIRAYADLAIVNLTDPERK 480
DB 421 klsaqrtqalattaisplshstshoballdfqaaklecepiirayadlaivnltdperk 480
OY 481 RSLHDYAKKLLQIOTFLFDFENORPHSPMYRLAYOVPESEFTKIMDLTEFLASKSESD 540
DB 481 rslhdyakllqiotflfddfennorphsmpyrlayovpeseftkimdlteflaskssed 540
OY 541 IRLLIQMTGEDAKNPFVLAGLITRIIVE 568
DB 541 irlliqmtgedaknfpvlagllitriive 568

```

RESULT 2  
AAV34620  
ID AAV34620 standard; Protein; 348 AA.

XX AAV34620;

XX 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae lipoprotein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

XX 03-JUN-1999.

PD 20-NOV-1998; 98WO-1B01890.

XX 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GENSET ) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

PS Page 640; Disclosure: 1912pp; English.

CC AAV34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.

XX Sequence 348 AA;

Query Match 57.1%; Score 1619; DB 20; Length 348;

Best Local Similarity 100.0%; Pred. No. 1e-126; Mismatches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MGLFHLTLFGLLCSLPISLVAKPEESVGHKILYISTOSTOQALATYLEALDAYGDHDF 60
DB 21 mglfhltlflllcslpislvakfpesvghkilyistgstqalatyaleadaygdhdf 80
OY 61 VLKRGEDYLNKSHSSDPQTRKSTITIGAGLSSEALDVLVSQAMETADPLOQLVLSAV 120
DB 81 vlkrigedylnkshssdpqtrkstltaglsgsealdvlsqametaadpqlqlvlsav 140
OY 121 SGHGLKTSDDLFFKALASPYVIRLEAAYRLANKTKFVIDHLSFTHKLPREIQCISAA 180
DB 141 sghglktsddlffkalsapyvirleaaayrlanlknkvidhlsfthklpreeicisaa 200
OY 181 IFLRLETESDAYIRDLAARSAIRSATALQIGEYQOKRFLPRLNLTSSASPODOEAI 240
DB 201 iflrleteesdayirldlaarsaaisrsatalqigeyqokrflprlnltsaspdodeai 260
OY 241 LVALGKLLKDGSGSYNNIKKQLOKPDVDTLAAQAALIALGKEEDALPVYIKKQALEERPRAL 300
DB 261 lvalgkllkdsgsyyinnikkqlokpdydtltaaqaalialgkeedalpvyikkqaleerpral 320
OY 301 YALRHLPESEIGIPALPIFLTKNSEA 327
DB 321 yalrhlpseigipalpifltknsea 347

```

RESULT 3

AAV34619	ID	AAV34619 standard; Protein: 245 AA.
XX	AC	AAV34619;
XX	DT	13-SEP-1999 (first entry)
XX	DE	Chlamydia pneumoniae transmembrane protein sequence.
XX	KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis
KW	KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW	KW	vaccine; neutralising epitope.
OS	OS	Chlamydia pneumoniae.
XX	PN	MO9927105-A2.
XX	PD	03-JUN-1999.
XX	PF	20-NOV-1998; 98WC-IB01890.
XX	PR	04-NOV-1998; 98US-0107078.
PR	PR	21-NOV-1997; 97FR-0014673.
XX	PA	(GENST ) GENSET.
XX	P1	Griffais R;
XX	DR	WPI; 1999-357842/30.
XX	PT	Genome sequence of Chlamydia pneumoniae
PS	PS	Page 639-640; Disclosure: 1912pp; English.
XX	XX	AAV34584-Y35879 represent the proteins encoded by all the open reading
CC	CC	frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC	CC	C. pneumoniae causes respiratory disease such as pneumonia and
CC	CC	bronchitis and is thought to be a contributing factor in heart
CC	CC	disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC	CC	nodosum or pharyngitis. The polypeptides encoded by the open reading
CC	CC	frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in
CC	CC	immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC	CC	nucleotides sequences can also be used as immunogenic compositions,
CC	CC	especially where the vector directs the expression of a neutralising
CC	CC	epitope of C. pneumoniae.
XX	Sequence	245 AA;

Query Match	42.9%	Score 1215:	DB 20:	Length 245:
Best Local Similarity	99.6%	Pred. No.3.2e-93:		
Matches 241:	Conservative 1:	Mismatches 0:	Indels 0:	Gaps 0:
QY	327	AKLVNALALLETGCDPKLLEYETRYTERLVOPHYENHLSFSKGRTLQONKRVNIIVPDP	386	
Db	4	sklnvalalileigcdtpkrlleyiterlvqphynelatsfskgrtlqwmkxvnlivpqp	63	
QY	387	QERELLSTTRGLEOIIITFLRPLKKEVYPCITVYKLASOKQQLTWTATISFISHSHQEA	446	
Db	64	qerellsttrgleeqilfltrlpkeaylpcitvylqlasqkqlatclatsishshqea	123	
QY	447	LDLLEQAAKLPEPILIRAYADAIYNLTRDPKKRSLDHYAKKLLQETLLFVDTENORPH	506	
Db	124	ldlllfqaaklppepilirayadaiaynltcdpekkrsldhyakkllqetllfvdtengrph	183	
QY	507	PSMPLRQVNPPESTKTLMDLTLEFLATSSSESDIRLLIQMTBEDAKNFVLAGILTKI	566	
Db	184	psmplirvqvpesctklmldtleclatsessedirllilqmtceqaknfpvlagiltki	243	
QY	567	VE 568		
Db	244	ve 245		

RESULT 4  
 AAY36830  
 ID AAY36830 standard; Protein; 468 AA.  
 XX  
 AC AAY36830;  
 XX  
 DT 07-OCT-1999 (first entry)  
 XX  
 DE Chlamydia trachomatis lipoprotein sequence.  
 XX  
 KW vaccine; eye disease; conventional trachoma; nongonococcal trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; peritrapatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
 KW  
 OS Chlamydia trachomatis.  
 XX  
 PN MO9928475-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98WO-IB01939.  
 XX  
 PR 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffais R;  
 XX  
 DR WPI; 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 PS  
 PS Disclosure; Page 708-710; 1755pp; English.  
 XX  
 AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 of Chlamydia trachomatis (see AAY01425). The polypeptides can be used as  
 vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 can also be used to control growth of the microorganism. Chlamydia  
 trachomatis is responsible for a large number of diseases, e.g. eye  
 diseases such as conventional trachoma, nongonococcal trachoma,  
 paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 peritrapatitis, Bartholinitis; pneumopathy in breast feeding infants;  
 and venereal lymphogranulomatosis. The polypeptides of the invention  
 may be of use in treating these diseases.  
 XX  
 Sequence 468 AA:  
 XX

	Query Match	Similarity	38.5%	Score 1092;	DB 20;	Length 468;
	Best Local	Similarity	51.5%;	Pred. No. 1.5e-82;		
	Matches	238;	Conservative	67;	Mismatches 135;	Indels 2; Gaps 1;
QY	1	MGLFHLTFGLGLCLPLSLVAKFPPEVSGHKLLTYISTOSTQCALATYLEALDAGDHDF	60			
DB	4	mglslatrlatfgfl--stflsaacdffpsvsqrrllfscqkspqpalaeayleavstlyqhdts	61			
QY	61	VLKRTGEDYLKQSHSSDPQTRKSTIIIGAGLGSSEDLVLSQAMETADPQQLLVLSNV	120			
DB	62	vlrvlaeaylyqgsflstcdthlrksaallqaglsyssealelllseatetqdllyeqalllnaa	121			
QY	121	SGHLKSTSDLLFLKALASPVIYRIEAAVRLANLNKTKYIDHLSFIIHKLPEIQCLSA	180			
DB	122	tsqjstksdskllfkyltsashpyrlaeayrilaacmknskvsdylyefiylpceiqnlaet	181			
QY	181	IFLRLETEBSDAYIRDLLAAKKSALRSATADLIGEYQQRFLPTLRNLNTLSASPQDEAI	240			
DB	182	iflqleteeadydyllrlsspnllcrnyvaaylyigevkqrrflprrlrslltsspsldqega	241			





OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li FMD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL15239.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS  
 PS Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC  
 CC  
 CC Sequence 7201 AA:  
 SQ  
 Query Match 4.6%; Score 129.5; DB 22; Length 7201;  
 Best Local Similarity 22.2%; Pred. No. 1.5;  
 Matches 135; Conservative 78; Mismatches 217; Indels 177; Gaps 28;  
 QY 41 QGALATYIEALDAYGDHDFVLRKIGEDYIKQSIHS-SDPQTRKSTTIGAGLAGSSEALD 99  
 DB 4991 egevaallqeeafdaye-----aln-k-akdyevglvkwdsdydq-----yteale 5034  
 QY 100 VLSQA---METADPLQQL-----VLSAVSGHL-----GRTSDDLFKA----- 135  
 DB 5035 wlskteaavgynkldgslkqkvvlegfgghlqtlfkwqktdldlmkagvllletcsdt 5094  
 QY 136 -----LASPYVI-----RLEAAYRLANLKNTKYIDHLHSFIRKLPEEI-OC- 176  
 DB 5095 risnaqlmqltkynaalltlakevmrtembyq-ehqghhslyeocgwsktreklesce 5153  
 QY 177 -----LSAATFLRIETESDAYIRPDLAAKSAIRSAITAOIGTYOQKRRLPT 224  
 DB 5154 qipgtlnevgklnvtvknltgfgtgnkltlyllelkekvimnteqnaaklqgd----- 5208  
 QY 225 LRNLTLSPDODRAIITAYALGKLMDGOSYVNIKKOLKPDVDTVLAAQAALIAIGK----- 280  
 DB 5209 -----tealkgdfkllvldln-----dvtrqklian-----rlaqleeifklykllle 5249  
 QY 281 -EEDALFVIRK-----QALEERPRALYALRHLRPEIG--IPALDIFLTKRSEAKLN 330  
 DB 5250 wledvepsvksrdeflndlsekraalekfrvqrdngndivakngrlkednsldldkd 5309  
 QY 331 VALALL---ELGCTPKLLEITRRLVOPH-----YNETIALSFSKGRITLQNN-KRVNI 380  
 DB 5310 fgpqlltkfdldqitgvnkllfesi-engvsnkkykqayne-----lqdwlrtrtri 5357  
 QY 381 IVPDDPERERRLSTGTGLEQILTFLEFRLPKKAVYLPCTIKYKLASOKTOLATITISFLSH 440

DB 5358 eweq-----cadehnekdyvesrlnrlgdigsslegkallaeceajsgavia----- 5405  
 QY 441 TSHQBALDLFQAKLPGEPIIRAYADLAIVMLTKDPKKRSIADYAKKLIQETL----- 495  
 DB 5406 tsqsegqdnvagekl-----hltsewelqgtsrdarslslesclawgt 5449  
 QY 496 -----LEFVDTENQRPHPSPYLRQVPPESRKLMLDLLEPLRLATSKSSDRLRLIQ 546  
 DB 5450 flqkfnklnlvtetmkrvtksqe--genktpedlvnakkllaevlaekndvedlnduce 5507  
 QY 547 LMTGDA 553  
 DB 5508 lmeqsa 5514  
 RESULT 9  
 ABG21017  
 ID ABG21017 standard; Protein; 1870 AA.  
 XX  
 AC ABG21017;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #21008.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS85204.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS  
 PS Claim 20; SEQ ID NO 51376; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1870 AA;

Query Match 4.5%; Score 128.5; DB 22; Length 1870;  
Best Local Similarity 19.2%; Pred. No. 0.27;  
Matches 102; Conservative 85; Mismatches 154; Indels 189; Gaps 26;

OY 141 PVIRLEAARLANLKNKTYIDHSHFIHKLPEE-----IQCLSAIFRLRE--T 187  
DB 1341 pvve-qalygfslnsk--sflinhihtlempesafarakvyfaalilvalhglykleyt 1397  
OY 188 EESDAYIRDL---AAKKSAL---RSATALOIGEQOKRFLPTLRNLTLASAPQOEAL 240  
DB 1398 dlnhtftellegvyvakpknmlrrsvle-----rmlsnwmsi-----c 1439  
OY 241 LVALGKLKDGQSYNN---IKKOLQKPDVDVTLAAQAALIALGKEEDALPVIRKQALEER 296  
DB 1440 lygylkdsagepilykifkalkhgvkgydav-----q 1472  
OY 297 PRALVALRHLRHPSEIGI-----PIALPIFLTKNSEAKLNYALALLEGCTPKILEY 348  
DB 1473 kkaeycl---ndtgllgdvayapltvsvlyvgdegvd---alpvkvlncdt---lsq 1520  
OY 349 ITERLVQPHYN-----ETIALSFSKGRFLQ-----NMKRVNIIVPOD 385  
DB 1521 vkekiidqvyrgqscwcpdpdsvlewprgstaqllsdldtsqregrtwkrvnlmhn 1580  
OY 386 POKERELKLTTRGLEQILTLFLRLPREAYLPC-----IYKLL-----ASOK 427  
DB 1581 vrdgatllskvygvsqgedsqdldpgerhalleeenvwhlvprcdevdegkskrgsvk 1640  
OY 428 TOLATTAISLSHT-----SHOEALDLLEFOAKLPGE---PLIRAYAD----- 467  
DB 1641 ekertkailteilylrlslsvkgtlqgfvdnfigsvlapghavppavkyftifldegaekhn 1700  
OY 468 -----LAIR-----NLTKDPEKKRSLH---DYAKKLIQETLLFVDTENORP 505  
DB 1701 igdedthlwktkslprlfrvwnllkphffldvhnveavdaslsvalgt---fmdacrite 1758  
OY 506 HPSMPYLKQVTPESRRTKMLDLETLANSKSSD-----IRLLIQLMTEG 551  
DB 1759 h-----klrdspsnkllyakeistykkmwedyykgyrrgmvyqsdg 1800  
RESULT 10  
ABG06301  
ID ABG06301 standard; Protein; 2017 AA.  
XX  
AC ABG06301;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #6292.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PT 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSR-) HYSO INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.  
DR N-PSTDB; AAG70488.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID No 36660; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 2017 AA;  
Query Match 4.5%; Score 127.5; DB 22; Length 2017;  
Best Local Similarity 20.0%; Pred. No. 0.36;  
Matches 113; Conservative 88; Mismatches 206; Indels 157; Gaps 26;  
OY 16 LPISLVAKPPEVSHKLTISTOSTQOALATYLEALDAYDHDHFVLRKIGEDYLKOSIH 75  
DB 625 ldvfyvlllgsevalndhlhtkqlkedkvenlvdqlnksqesvnsqikenle--lkehxr 682  
OY 76 SSDPQ---TRKSTIIGGLAGSSSEALDVL--SQAMENADPLQQLVLSAVSGHGKTSDD 130  
DB 683 qneeeisrltnelmgslngdsnsnfkdtllkeraeavtnlknlseleqinenlkkvaf 742  
OY 131 LLEKALASPYVIRLEAAYRLANLKNKTYIDHSHFIHKLPEEIQCLSAIFRLETE-- 188  
DB 743 v-----xmenekilylacedvnhqlee---clagnnqlslentl 778  
OY 189 -----ESDAYIRDLAAKKSALRSATALOIGEQOK-RPLPLRLNLTLASAPQOEAL 241  
DB 779 vetlmekegleleaelcwakrrlllea-----nxyeklieelsnarnlntalsqlqehel 833  
OY 242 YALGKLKDGQSYNNIKKOLQKPDV-----VTLAAQAALIAL-GKEEDALPVYK-- 289  
DB 834 -klnqkd-metaelkknqemdcldhketdvjssleegkqitqllnkkeitlekxr 891  
OY 290 -----KQALERRPALYALR---HLPSEIGIPALPIFLTKN 324  
DB 892 sklgeelqysqalrknellrqtieekdtrslgmkeenhlgee-----lerlr 941  
OY 325 SEAKLNVALLLEIGCTPFLLEYITE-----RLVQPHVETLALSTSKGRITLONMK 376  
DB 942 eegsrtaeva-----dpktdlsvtelaseveqlnlkehlkeelk-hnqklliedqns 993  
OY 377 RVNIIVQDPOKERELKLTTRGLEQI---LTLFLRLPREAYLPCITKLLASOKTOLATY 433  
DB 994 kmqll--qslqeqkemdelfryqhegnathgtlf-lekdeeklsqktleqitql--- 1047  
OY 434 AISFLSHTSHOEALD-----LFOAKLPGEPILIRAYADLAIYNLTKDEKKRSLDVA 487

```

Db 1048 -----heerddqtdnsdftgctkxys-----lntengsek-----hals 10832
QY 488 K-----KLIQETLFFVDTENOR 504
1083 kaeterlvkikereleiklinek 1106

```

## RESULT 11

ID	AA99675 standard; Protein; 631 AA.
XX	
AC	AA99675;
XX	
DT	10-OCT-1996 (first entry)
XX	
DE	RHAMM 1-2a isoform.
XX	
KW	RHAMM 1-2a: receptor for hyaluronic acid mediated motility
KM	hyaluronan receptor; cell locomotion; cell proliferation;
KX	breast cancer; therapy.

PN	EP721012-A2.	
XX		
PD	10-JUL-1996.	
XX		
PF	16-OCT-1995;	95EP-0307310
XX		
PR	14-OCT-1994;	94GB-0020740

Claim 8; Page 50-52; 117pp; English.

RHAMM 1-2a (RNR99675) is an alternatively spliced variant of RHAMM 1 (RNR99673) (receptor for hyaluronic acid mediated motility), a protein involved in cell locomotion or motility and cell proliferation and transformation. It differs from RHAMM 1 by an insertion of 23 amino acids (see also RNR99674) between amino acids 54 and 55 of RHAMM 1, resulting from an alternatively spliced exon 2a (AAT34502). RHAMM 1-2a is the isoform that is overexpressed in tumours. Determination of the level of RHAMM 1-2a in a sample can be used to assess the prognosis of a tumour (esp. breast cancer) patient. The RHAMM 1-2a protein can also be used to suppress or control a tumour by modulating the interaction of cell-associated RHAMM with its ligand.

SQ Sequence 631 AA;

Query Match	4.48;	Score 126;	DB 17;	Length 631;
Best Local Similarity	21.38;	Pred. No. 0.091;		
Matches 115;	Conservative 97;	Mismatches 213;	Indels 114;	Gaps 23;

QY 70 LKSIHSSDPQTAKSTII---GAGIAGSSEALDVL SQAMETADPIQLQLLVLSAVSGHLKG 126  
|| | : : : || : : : || : : :  
Db 31 Iklgatgkdlteskgkivqlegk|vslekerd--ekceteklleyiqelscasadvkek 87

```

OY      127 TSDDLFFALASPVVILEAATRLANKRVNIDHLSFTIKU-?EEDIOCSAAIFLR 1840
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      88 kcvd-----iaqleedk---ektarellslglsleenitfskqledltvkcql- 1322
OY      185 LETESDAYIRD---LAAKKAISANVALDIGEV--QOKREPLTLRNLITTSASP0D 236
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      133 leterdnrvskdreanaeliseemqllteralajeregeyikqke-----lqsagll 184
OY      237 QEALUYA-LGKLMDGOSYVNIKKOLOPVDVTLLAQAOLIALGKEEDALPVYIKKALE 294
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      185 qekelsatlgqqalsfgsemtseknvikeekiala--eldavqgkeesgerilivqlee 2411
OY      295 ERPRALVALRHLPSIEIGIPALPIELTKNSEAKLNVA-----LALLELCDTPKLEY 348
Db      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      242 erkstaegltfrldn-----lirekevелеkhiaahagaqalllaqeykndtagrlrd 291
OY      349 ITERL-VQPHYNETLALSFGSKGTIONMKRVNIIIVQDOERRLSTTRGLEEQILTF 406
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      292 vtaqlasveqekyndt-----aqslrd----vtaqlasegeknyndtagslrld-vta 336
OY      407 LFRLPKEAY-----LEPCIYKLASOKTOLATTAFISFLSTSHOEADULLFOAKLPGEP 460
Db      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
      337 qlasegeknyndtagslrldvtacqlasvegeknyndtagslrldvtaqlas----yksstl---- 388
OY      461 IIRAYADLAIVNLTKDPKRSLHDVAKKLIQETLFLVTDENORPHSPMYLRYOYTPEL 520
Db      || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      389 --ketedliklenltl--gekvamaekevdevgqlllaestinge-----yarmvqd|qp 438
OY      521 RTKMLDLIELLATL-----KASEDRILLIOMTEGDGAKNFYPVAGLLIKI 566
Db      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      439 rntlkeeiektusflekittidklngqlrgdqedrtfqyleeqkrteekenvmelmelmeel 497

```

**RESULT 12**

ID	AA	99673	standard;	Protein;	606	AA.
...						

AC	AAR99673;
XX	
DT	10-OCT-1996 (first entry)
XX	

DE Receptor for hyaluronic acid mediated motility RHAMM 1.  
XX  
KM RHAMM 1; receptor for hyaluronic acid mediated motility;  
KM hyaluronan receptor; cell locomotion; cell proliferation  
KM breast cancer; therapy.

Mus sp. OS

FT	Key	Location/Qualifiers
FT	Modified-site	91..93
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	258..260
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	279..281
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	300..302
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	321..323
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	342..344
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	373..375
FT	Modified-site	/label= N-glycosylation_site
FT	Modified-site	413..415
FT	Modified-site	/label= N-glycosylation_site
FT	Domain	532..542
FT	Domain	/label= Hyaluronan_binding_domain-I
FT	Domain	553..562
FT	Domain	/label= Hyaluronan_binding_domain-I
FT	Modified-site	594..596
FT	Modified-site	/label= N-glycosylation_site





OY		70	LKQSHSSDPQRSTII---GAGIASSSEALDVLQSAMETADPDLQOLLVLSAVSGHLK	126
Db		31	ILlqgtqkdlteskykylvlgelkvlslekeId---eketeklleylgelscasdqvsk	87
OY		127	TSDDLLFFALASPYPVIRLEAAVYRLANKTKRVVDHLSFTIKLV--PEEIOCSAIFR	184
Db		88	kcvd-----laqlaedlk---ektdeislksyleemltfskgledltvkcgll-132	
OY		185	LETESDAVIYND----LLAKKSAINSAVALDIGEV---OQRFLPTLRNLNLTSPD	236
Db		133	Ieterdnivskdreraeellsaemqilferlalergeryeklqke-----lgsqslly	184
OY		237	GEALITYA--LCKMLDGOSYNIKKQQLDKPDVDYLAAAOGLINLKGEDALPIYKQALE	294
Db		185	gekelasrlggqlsgfgeemtseknvtkeeklala---eldavvgkeeqserlvkglee	241
OY		295	EPRPALVALRHLPSREIGIPALPIFLTKNSEAKLNVA-----LALLEGCPTPKLEY	348
Db		242	erkstaegltrlldn-----llrekevelekhhuaagalllagekyndtaqsird	291
OY		349	ITERL--VOPIRYNETLASFSKGFTLRNMKRNVITVPQDERELLSTTRGLEOILNF	406
Db		292	vtaqlasvgekynnt-----agslrd-----vtaqlasegekynndaqsird-vla	336
OY		407	LFRLPKEY-----LPCIVKLASOKTQOLATTAISLTSHQEALDDLFOAKLPGEP	460
Db		337	qlasegekynndaqsirldvtaqlasge-kyndtaqsirldvtaqlas---yksstl---	387
OY		461	IRRAYADLAIVNLKDEPKRRSLIDVAKKILIOETLFDVTENGRPHPSMPLRYGVTPES	520
Db		388	--keledtklanhlcl--gekvanmeksvedvqqilltaestnge-----yarwvqdldgn	437
OY		521	RTKMLDLIEFLANS-----KSEEDIRLLIQLMTEGDARKNPVLAGLDIKI	566
Db		438	rslakeekeltessfleklrldlnqrqgdgedfrkxgleekyrtaekenmtlmeel	496
<hr/>				
RESULT 14				
ID	AAU00015	standard; Protein; 1838 AA.		
XX	AAU00015;			
DT	09-MAY-2001	(first entry)		
XX				
DE	Human Plexin-B2.			
XX				
KW	Human: Plexin-B2; semaphorin domain; hyperplasia; neoplasia; cancer;			
KW	neurodegenerative disease; autoimmune disease; lupus; multiple sclerosis;			
KW	Inflammatory bowel disease; diabetes type I; rheumatoid arthritis;			
XX	Immunogen; antibody; MET related sequence; MRS.			
XX				
DS	Homo sapiens.			
XX				
FT	Key	Location/Qualifiers		
FT	Peptide	1..18		
FT	/label=	Signal_peptide		
FT	Domain	1..1199		
FT	/label=	"Extracellular domain"		
FT	Protein	19..1838		
FT	/label=	"Mature Plexin-B2"		
FT	Domain	19..518		
FT	/label=	"Semaphorin domain"		
FT	Region	451..530		
FT	/label=	"Primary Met related sequence, MRS"		
FT	Region	601..680		
FT	/label=	"Secondary MET related sequence, MRS"		
FT	Region	751..830		
FT	/label=	"Tertiary MET related sequence, MRS"		
FT	Region	800..1010		
FT	/label=	"G-P repeat region"		
FT	Domain	1196..1215		
FT	/label=	"Transmembrane domain"		

XX	PX	MO200114420-AZ.
XX	PD	01-MAR-2001.
XX	PF	25-AUG-2000; 200OWO-US23365.
XX	PR	25-AUG-1999; 99US-0150576.
XX	PA	(UYTO-) UNIV TORINO.
XX	PA	(REGC ) UNIV CALIFORNIA.
PI	PI	Artigiani S, Comoglio PM, Goodman CS, Tester-Lavigne M;
PI	PI	Tamagnone L;
DR	DR	WPI: 2001-226610/23.
XX	XX	N-PSDB: AAS00020.
PT	PT	New plexin polynucleotides and polypeptides, useful in diagnosis,
PT	PT	therapy and in producing compounds for treating diseases involving
PT	PT	aberrant cell growth (e.g. cancer) or immune regulation (e.g.
PS	PS	autoimmune diseases)
PS	PS	Claim 4; Page 52-56; 79pp; English.
CC	CC	The sequence represents Human Plexin-B2. Plexins are large transmembrane
CC	CC	proteins whose extracellular domain shares homology with Scatter factor
CC	CC	receptors and contain an approximately 500 amino acid Semaphorin domain.
CC	CC	The plexin polynucleotides and polypeptides, and plexin-specific binding
CC	CC	agents are useful in diagnosis, therapy and in the pharmaceutical
CC	CC	industry. In particular, the plexin polynucleotides and polypeptides are
CC	CC	useful for generating compounds (e.g. plexin-specific binding agents or
CC	CC	antibodies) for treating or diagnosing a disease or disorder involving
CC	CC	aberrant cell growth (e.g. hyperplasia, neoplasia, cancer or
CC	CC	neurodegenerative disease), or diseases or disorders involving aberrant
CC	CC	immune regulation (e.g. autoimmune diseases such as lupus, inflammatory
CC	CC	bowel disease or diabetes Type I), or immunosuppressive diseases such as
CC	CC	multiple sclerosis or rheumatoid arthritis.
XX	SQ	Sequence 1838 AA;
OY	OY	Query Match 4.3%; Score 121.5; DB 22; Length 1838;
OY	OY	Best Local Similarity 18.9%, Pred. No. 1.
OY	OY	Matches 98; Conservative 84; Mismatches 162; Indels 175; Gaps 23
Dy	Dy	141 PYTRLEAAVRLNLTNKTKYIDHLHSFIRHKLPPE-----IQCLSAIPLRE--T 187
Dy	Dy	1310 pyve-galyqfnnlnsk--sfllnfihlengretsarakyyfasillvalngkleyt 1366
Oy	Oy	188 EESDAYIRDLLAAKSKAIRSATALOIGEXYOQRFLPTELRNLTLTSASPODEALVALGKL 247
Dy	Dy	1367 dlmhtclfiellleegyvaknpklmlrtselvermismnmsi-----clyylkd 1415
Oy	Oy	248 KDGOSTYN----IKQLOKRPDVTVLAAQAOLIALGKEEDALPVYKKQALEERPALVYL 303
Dy	Dy	1416 sagepllylkfkaihqvekypvdav-----gkkakylt 1448
Oy	Oy	304 RHLPEIGI-----PALPIPLFKTNSEAKLWVALALLELGCDTPKLBETIPFRWLQ 355
Dy	Dy	1449 -----ndvltlgddveyaprlvsivtygdvyd---alpvkvlnct--lsqvkekild 1496
Oy	Oy	356 PHYN-----FTLALSFSGKGRITQ-----NMKRNIITVPDDPOERERL 392
Dy	Dy	1497 qvryrgqpcscwprpsavllepwpstqaqlldsdldtsqregtwkrvntlmhyrvrdgatc 1556
Oy	Oy	393 LSTRGLEEQILTELFRLLPKAYLPC-----LYKLL-----ASOKTOLATTA 434
Dy	Dy	1557 llskvysvqpedsqdqlpgerhalleenrwhlvrrptdevdegkskrsvkekerka 1616
Oy	Oy	435 ISFLSHT-----SHQEAIDLFPAAKLPGE---PIIRAVD-----L 468
Dy	Dy	1617 itelyltiltlsvkgllqgvfdvnftfsvglapgnhavppakkyffidfdgaekhnldedti 1676

YY 469 AIV-----NLTDPEKKRSRH-----DYAKRLTLETFVDTNORHPSPMYL 512  
Db 1677 hmwknsuplrfvwiilknphlfdvhnvevdasasvtaqf--fmdactrtth----- 1728  
QY 513 RYQVTPESRPTKIMLDILETFLATSKSED---IRLLIQL 547  
Db 1729 --klardspnskllYakeistykkmvedyykgirgmvyv 1765  
RESULT 15  
ABBI1818  
ID ABBI1818 standard: peptide; 1849 AA.  
XX  
AC ABBI1818;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human plexin-B1/SEP receptor homologue, SEQ ID NO:2188.  
XX  
KW Human: cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cystostatic; osteopathic; vasotropic; cardiant; virocidic; antibacterial;  
KW antifungal; vulnerary; antitumor.  
XX  
XX Homo sapiens.  
OS  
PN WO200157188-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 05-FEB-2001; 2001WO-US03800.  
XX  
PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Dirmannac RT;  
XX  
DR WPI; 2001-457740/49.  
DR N-PSDB; ABA09062.  
XX  
PT Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
XX e.g. arthritis and cancer -  
XX  
PS Claim 20; Page 259-260; 1963pp; English.  
XX  
CC Sequences ABBI0991-ABBI1330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or

CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention.

	Query Match	4.3%;	Score 121.5; DB 22; Length 1849;
	Best Local Similarity	18.9%; Pred. No. 1;	
	Matches	98; Conservative	84; Mismatches 162; Indels 175; Gaps 23;
QY	141 PVIRLEAAVRLANLKNTRVIDHLNHSFIKLPEE-----IQCSAIFLRLE-T	187	
Dd	1321 pve-galyqfennlnsk-sflnfhltlenqsfesarakyyfasiltvalhygleyt	1377	
OY	188 EESDAVIINDLLAAKSAIRSATADIGEQKRFLPTLRNLITSSAPDQEALYALGKL	247	
Dd	1378 dmhtlfllelegvvaknpkmlirsetveermIsnmwsl-----cLyqykld	1426	
OY	248 KDGOSYVN---IKKOLCKPDVDVTLMQAQALIALGKEBDALPVIAKKOAEERPAVAL	303	
Dd	1427 sagepllylkfkaihqekyvday-----qkakkytl	1459	
OY	304 RHLPSEIGI-----PIALPILFRTKNSSEAKLNVALALELGCDTPKLEYITERLWO	355	
Dd	1460 ----ndtgllggdvveyaprlvtvsivlgdegvd----alpvkrklncvtt--lsqvkekild	1507	
OY	356 PHYN-----ETLASFSKGTLQ-----NMKANIIIVDDPODERREL	392	
Dd	1508 qyryqpccscwprrpsavlwewrpgstaqlsqldltsqgregtwkwvntlmhnvudgatll	1567	
OY	393 LSTNGLEGQLTFLFRLPKEAYLPC-----TYKLL-----ASOKTOLATTA	434	
Dd	1568 llskvgvsqgedsqddlpgerhallleenrynwhlvrtptdevdegkskstrsvkekerka	1627	
OY	435 ISFLSHT-----SHOEALDLDFQAANKPGC---PIIRAVAD-----L	468	
Dd	1628 lteilyltrllswxkgylqgfvdnfigsyapghayvparkyfiddfldegaekmidedeti	1687	
OY	469 AIY-----NLTKDEPKRSLH-----DYAKKLIQETLLFDVTENORHPSPMYL	512	
Dd	1688 hmwktnsiplrftwnllknphflifdvhvnevdaslasvaqt- fmdactreth-----	1739	
OY	513 RYOVPBESTKMLMDILETLATSSESQD-----IRILLQL	547	
Dd	1740 --klardspsnklyakeistyxkhnevdyxyglrqmvqu	1776	

Search completed: August 20, 2002, 04:10:13  
Job time: 3483 sec

---



```

QY 31 KILYISTOST-QQALATYLEALDAYGDHDFVLRKIGEDYLKQSIHSDPQTR----- 82
Db 1107 KVLGETOSSLMKAVAFLEST---AMHDIMA---AEKYFGTATGNRPSQEGERYNY 1159
QY 83 -KSTIIAGLAGSSBALDVLISQAMETADPLQDLVLISAVSHGLGTSDDLFKALASPPY 141
Db 1160 SKCTIV-----VRIEMTTTLSTSPGKMLLEKDVONTMLM-----LVKTKICEPSS 1208
QY 142 VIRLEAARLANLKNTKYIDHLHSFTIHKLPETIQLSAIFLRLETESDAYIRDLAAK 201
Db 1209 I-----GFNIGDVAVMNYPVCTNL-----MKALK 1234
QY 202 KSAIRSATALOIGEYQQRFLPTLRNLTSASPODEAI-LYALGKLDG---QSYYNK 257
Db 1235 KSPYDILIEH-----LKEKITAQSIIEELCAVDLYCPACVDRARLASVNSAC 1282
QY 258 KOLQRPDYVTLAAQAAL-----IALGKEBALPVI---KQQA--LE 294
Db 1283 KOLHRAGVLCVITIPQSADQHHISIGTKLISLYKSIAPGDEQCCPLSDPNCKRLASGL 1342
QY 295 ERPRALYAL-RHLPSET--GIPIALP-----IFLKTNSSEAKLNV 331
Db 1343 ELAFAFGGLCEHLVSLDDTVLSPSRGSGQKNIVSFSGEYFSLPSETINTELLKNL 1402
QY 332 ALALIEL---GCDPCKILEYI-----TERLVQPHYNETLALSFSGKRTLQWKKRYN 379
Db 1403 DLAVIELMKSSVDNPKMYSNVNGMLDOSFRDRSEKHQGLKLAIT---LQWKKCD 1457
QY 380 IIVPDPQERERLLSTTRGLEBOILTFRLPKREAYL--PCIYKLLASOKTOLATTAIS 436
Db 1458 SWMAKDSAPESKMAVLT-----LLAKIFQIDSSVCFTNMCHM-----PEVETTVYS 1504
QY 437 FLSHSHQEAIDLFLQAALKGEPITIRAYADLAIT-NITKPEKKRSLHDYAKKLIQETL 495
Db 1505 LLAOSK---IDL-----HLGQALITL---LPFTSLTGG---SLEDL--KVVLN 1544
QY 496 L---FVDTENORPHSPMYLRYQVTPESRTKIMLDILETLATSKSSEDIRLLIOMTE 550
Db 1545 IVSNFPMKSEFPPTGTLQYNNY-----VDCMKKFLDALELSPP---MLLOLMT 1591

RESULT 2
US-09-407-562-29
: Sequence 29, Application US/09407562
: Patent No. 6294334
: GENERAL INFORMATION:
: APPLICANT: Kathryn Week
: TITLE OF INVENTION: Genetic Test For Equine Severe
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dr. Benjamin A. Adler
: STREET: 8011 Candle Lane
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: Macintosh
: SOFTWARE: Microsoft word for Macintosh
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/407,562
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/970,269
: FILING DATE: NO. 6294334ember 14, 1997
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Adler Ph.D., Benjamin A.
: REGISTRATION NUMBER: 35,423

```

```

: REFERENCE/DOCKET NUMBER: D5860
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 713-777-2321
: TELEFAX: 713-777-6908
: INFORMATION FOR SEQ ID NO: 29:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2987 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: no
: ANTI-SENSE: no
: ORIGINAL SOURCE:
: FEATURE:
: US-09-407-562-29

Query Match 4.3%; Score 121.5; DB 4; Length 2987;
Best Local Similarity 21.4%; Pred. No. 0.16;
Matches 128; Conservative 83; Mismatches 196; Indels 191; Gaps 32;

QY 31 KILYISTOST-QQALATYLEALDAYGDHDFVLRKIGEDYLKQSIHSDPQTR----- 82
Db 1107 KVLGETOSSLMKAVAFLEST---AMHDIMA---AEKYFGTATGNRPSQEGERYNY 1159
QY 83 -KSTIIAGLAGSSBALDVLISQAMETADPLQDLVLISAVSHGLGTSDDLFKALASPPY 141
Db 1160 SKCTIV-----VRIEMTTTLSTSPGKMLLEKDVONTMLM-----LVKTKICEPSS 1208
QY 142 VIRLEAARLANLKNTKYIDHLHSFTIHKLPETIQLSAIFLRLETESDAYIRDLAAK 201
Db 1209 I-----GFNIGDVAVMNYPVCTNL-----MKALK 1234
QY 202 KSAIRSATALOIGEYQQRFLPTLRNLTSASPODEAI-LYALGKLDG---QSYYNK 257
Db 1235 KSPYDILIEH-----LKEKITAQSIIEELCAVDLYCPACVDRARLASVNSAC 1282
QY 258 KOLQRPDYVTLAAQAAL-----IALGKEBALPVI---KQQA--LE 294
Db 1283 KOLHRAGVLCVITIPQSADQHHISIGTKLISLYKSIAPGDEQCCPLSDPNCKRLASGL 1342
QY 295 ERPRALYAL-RHLPSET--GIPIALP-----IFLKTNSSEAKLNV 331
Db 1343 ELAFAFGGLCEHLVSLDDTVLSPSRGSGQKNIVSFSGEYFSLPSETINTELLKNL 1402
QY 332 ALALIEL---GCDPCKILEYI-----TERLVQPHYNETLALSFSGKRTLQWKKRYN 379
Db 1403 DLAVIELMKSSVDNPKMYSNVNGMLDOSFRDRSEKHQGLKLAIT---LQWKKCD 1457
QY 380 IIVPDPQERERLLSTTRGLEBOILTFRLPKREAYL--PCIYKLLASOKTOLATTAIS 436
Db 1458 SWMAKDSAPESKMAVLT-----LLAKIFQIDSSVCFTNMCHM-----PEVETTVYS 1504
QY 437 FLSHSHQEAIDLFLQAALKGEPITIRAYADLAIT-NITKPEKKRSLHDYAKKLIQETL 495
Db 1505 LLAOSK---IDL-----HLGQALITL---LPFTSLTGG---SLEDL--KVVLN 1544
QY 496 L---FVDTENORPHSPMYLRYQVTPESRTKIMLDILETLATSKSSEDIRLLIOMTE 550
Db 1545 IVSNFPMKSEFPPTGTLQYNNY-----VDCMKKFLDALELSPP---MLLOLMT 1591

RESULT 3
US-08-970-269A-30
: Sequence 30, Application US/08970269A
: Patent No. 5976803
: GENERAL INFORMATION:
: APPLICANT: Kathryn Week
: TITLE OF INVENTION: Genetic Test For Equine Severe
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:

```







```

: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/195,487
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/901,701
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7000
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-195-487-4

```

Query Match 4.0%; Score 113; DB 1; Length 2101;

Best Local Similarity 21.9%; Pred. No. 0.56; Mismatches 184; Indels 138; Gaps 25;

```

: 38 QSTQOALATYIEALDAYDHFVLRKIGEDY-LKQSIHSDPQTRKSTIGAGIAGSSE 96
: 1387 QSKOAGGLRAELRLAQ-----RELGLIPLRQKVAED---RTAQQLRAEKASYAE 1435
:
: 97 ALDVLQAMETADPLQQLLVLSAVSGHLGKTS-----DLFKALASPPVIRLE 146
: 1436 QLSMLKKRHG-----LAEENRGIGERANLGRQFLEVLDQAREKYYQELAAVRAD 1486
:
: 147 AAYRLANLK-----NTRVIDHLSFTHKLPPEETQCSAAIFLRLE 186
: 1487 AETRLAEVQRAOSTARELEVMTAKYEGAKVKVLEERQF---QEEROKLTA-----Q 1536
:
: 187 TEESDAYIRDLAAKKSAIRSATAIOT--GEYQO--KRFPLTRLNLTSASPODEALTY 242
: 1537 VEELSKRLADSDQASKYQOQKLKAVQAGGESQDEAQRFOQLNELQALSQKEQAABHY 1596
:
: 243 ALGKLKDGOSYYNIKK-----OLQRPDVDTLAAQAOLALGKEEDALFVI 288
: 1597 KL-QMERKAKTHYDAKKQOQNOELOERLSELOKRNKELRAEAER---LGHEIQQAQGLK 1651
:
: 289 KKQALEERPRLALYA-LRHLPSIEIG-----IPALPLFLTKNSEAK--LNVALA 334
: 1652 TKEA-EQTCRHILTAQVRSLAEQAVAHADQQLRDLGKFQVATDA-LKSREPQAKPOLDLSID 1709
:
: 335 LLEIGCDPRLKLEYTEITTELVQPHYNETLALFSKGRITLQNMKRVNIIVPOPOE--RRRL 392
: 1710 SLDSLCEGTPLS-ITSKLPTQPDGT-----SYGPEPASISRL 1749
:
: 393 LSTRGLEEQILTFV---FRLPKEAYLPCIVKLL--ASQKTQALATTAISFHSHTSHQAL 447
:
: 448 DLFLQAAKLPGEPITIRADIAIYNLTDPPEKKRSL 483
:
: 1750 PPKVESLESLYFTPIPARSQAPLESLSIDSLDVFIDSGRKTRRSARRPTQIINIMTKKL 1809
:
: 1810 DV-----EPP--DSANSSFYSTRSPASQASL 1834

```

RESULT 7  
PCT-US93-06160-4

; Sequence 4, Application PC/TUS9306160  
; GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE

TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: TESTA HUMWITZ & THIBEAULT  
STREET: 53 STATE STREET

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/06160

FILING DATE: 19930621

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: PITCHER ESQ, EDMUND R

REGISTRATION NUMBER: 27,829

REFERENCE/DOCKET NUMBER: MTP-013

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/248-7000

TELEFAX: 617/248-7100

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 2101 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US93-06160-4

Query Match 4.0%; Score 113; DB 5; Length 2101;

Best Local Similarity 21.9%; Pred. No. 0.56; Mismatches 184; Indels 138; Gaps 25;

```

: 38 QSTQOALATYIEALDAYDHFVLRKIGEDY-LKQSIHSDPQTRKSTIGAGIAGSSE 96
: 1387 QSKOAGGLRAELRLAQ-----RELGLIPLRQKVAED---RTAQQLRAEKASYAE 1435
:
: 97 ALDVLQAMETADPLQQLLVLSAVSGHLGKTS-----DLFKALASPPVIRLE 146
: 1436 QLSMLKKRHG-----LAEENRGIGERANLGRQFLEVLDQAREKYYQELAAVRAD 1486
:
: 147 AAYRLANLK-----NTRVIDHLSFTHKLPPEETQCSAAIFLRLE 186
: 1487 AETRLAEVQRAOSTARELEVMTAKYEGAKVKVLEERQF---QEEROKLTA-----Q 1536
:
: 187 TEESDAYIRDLAAKKSAIRSATAIOT--GEYQO--KRFPLTRLNLTSASPODEALTY 242
: 1537 VEELSKRLADSDQASKYQOQKLKAVQAGGESQDEAQRFOQLNELQALSQKEQAABHY 1596
:
: 243 ALGKLKDGOSYYNIKK-----OLQRPDVDTLAAQAOLALGKEEDALFVI 288
: 1597 KL-QMERKAKTHYDAKKQOQNOELOERLSELOKRNKELRAEAER---LGHEIQQAQGLK 1651
:
: 289 KKQALEERPRLALYA-LRHLPSIEIG-----IPALPLFLTKNSEAK--LNVALA 334
: 1652 TKEA-EQTCRHILTAQVRSLAEQAVAHADQQLRDLGKFQVATDA-LKSREPQAKPOLDLSID 1709
:
: 335 LLEIGCDPRLKLEYTEITTELVQPHYNETLALFSKGRITLQNMKRVNIIVPOPOE--RRRL 392
: 1710 SLDSLCEGTPLS-ITSKLPTQPDGT-----SYGPEPASISRL 1749
:
: 393 LSTRGLEEQILTFV---FRLPKEAYLPCIVKLL--ASQKTQALATTAISFHSHTSHQAL 447

```

Db 1750 PKVESLESLYTPIPARQAPLESLSIDGVFLDSGKRTSARRRTTQIINITTKKL 1809  
QY 448 DLLEQAKLPEPIIRAYADLAITYNLTKDPEKKRSL 483  
Db 1810 DV-----EEP---DSANSSPFSRSPASQASL 1834

RESULT 8  
US-08-117-362-3  
; Sequence 3, Application US/08117362  
; Patent No. 5595872  
; GENERAL INFORMATION:  
; APPLICANT: Wettlau II, John R.  
; APPLICANT: Sharp, Daru Y.  
; APPLICANT: Gregg, Richard E.  
; TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burton Rodney  
; STREET: P.O. Box 4000  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08543-4000  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/117,362  
; FILING DATE: 03-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gaul, Timothy J.  
; REGISTRATION NUMBER: 33,111  
; REFERENCE/DOCKET NUMBER: DC21b  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252-4526  
; TELEFAX: (609) 252-5901  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 860 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-117-362-3

Query Match 4.0%; Score 112; DB 1; Length 860;  
Best Local Similarity 19.8%; Pred. No. 0.16;  
Matches 127; Conservative 98; Mismatches 226; Indels 192; Gaps 32;

QY 30 HKILYISTOSTQOALATYIEALDAYGDHDFYLRKIGEY-----LKSHSSDPQTRK 83  
Db 171 HOVLGVTSKAT-----SVTTYKIEDSEFVAVLSEIRALRNLFQSI----- 212

QY 84 STIIAGLAGSSEALDV-LSQAMETADPLQOOLLYSAVSGHLGKTSDDLFEALASPPV 142  
Db 213 -----AGKIVSRQKLEKTEESVRLKQKVAA-----IKAVDSKTYTA 252

QY 143 IRLFAAYRLANKN-TKYIDLHSF-IHKLPEIQLSAA-IFLRLETEESDAYIRDLA 199  
Db 253 IPIVGQVFSQKCGPQSEHMQSIRKHLQPNLSKAEAVRSFL-----AFIKHLRT 304

QY 200 AKKSAIRSATALQIGEYQCKRFLPTLRNLLISASPOD-----QEAR 240  
Db 305 AKKEEI-----LQILKAKEKEVLPOLVDVAVTSAGTPDSLDAILDFDKSTESVILQERF 359

QY 241 LYA-----LGRKLDGOSYVNIKKQLQKDPVDVTLAA----- 271

Db 360 LYACAFASHPDEELRLALISKFKSGFSNDIRE-----SVMTIICALVYRKLCONOGCKIK 414  
QY 272 ----AQALIALGKEEDALPVIRKQALEERPRALYALRHLPSEIGIPALPIFLK-TKNSE 326  
Db 415 GVEAKKLLIGLGE-----KAEREKEDIVMYLLAKN-----ARLEGGIPLLLKYTEGE 463

QY 327 AKLNVALALLEGCCTPKLLEYITERLVQ-----HYNETL-----ALSFSGKRT 371  
Db 464 GPISHLAATTLQRYDVP-----FTTDEVKRTMNRITYHQNKRIHEKTVRTAAALITLKNPS 519

QY 372 LQMKRVNITIVPODERERLLSTTGLLEOILFLRLPKRAYLPCIKLLAS----- 425  
Db 520 YHEVKNILLISIGELPKEMKMYLS---IYQDILR--FERPASMVQVLUKEVAHNYDF 574

QY 426 QKTQATTAISPLSHTSHOE--ALDLFQAAKLPEPIIRAYADLAITYNLTKDPEKKRS 482  
Db 575 SKGSSSAVGYVERTSHASTYSDILYS-----GSGILRR-SNLNIFYI-----EKTP 624

QY 483 LHDYAKKLLQELT--LFVYT-----ENQRPHSPMYLR--QVTPESRTKMLDILETLA 533  
Db 625 LHGIQVIEAOGLEALIAATPDEGEENLDSYGLSALLPFDVQLRPVTFNGYSDLMSKML 684

QY 534 TSKSS-----EDIRLLQLMTE-----GDAKNFPVLAGLLIKI 566  
Db 685 SASDPMMSYKGLLLILIDHSQELQDSGLKANMDVGGGLAIDI 727

RESULT 9  
US-08-486-924-3  
; Sequence 3, Application US/08486924  
; Patent No. 5789197  
; GENERAL INFORMATION:  
; APPLICANT: Wettlau II, John R.  
; APPLICANT: Sharp, Daru Y.  
; APPLICANT: Gregg, Richard E.  
; TITLE OF INVENTION: MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burton Rodney  
; STREET: P.O. Box 4000  
; CITY: Princeton  
; STATE: New Jersey  
; COUNTRY: U.S.A.  
; ZIP: 08543-4000  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,924  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/117,362  
; FILING DATE: 03-SEP-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gaul, Timothy J.  
; REGISTRATION NUMBER: 33,111  
; REFERENCE/DOCKET NUMBER: DC21b  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 252-4526  
; TELEFAX: (609) 252-5901  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 860 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-486-924-3

Query Match 4.0%; Score 112; DB 1; Length 860;  
Best Local Similarity 19.8%; Pred. NO. 0.16;  
Matches 127; Conservative 98; Mismatches 226; Indels 192; Gaps 32;

```

QY 30 HKILISTOSTQOQLATYLELDAVGHDFVLRKIGEDV-----LKOSIHSDPQTR 83
Db 171 HOVLGVTSKAT-----SVTTYRKEDSFVAVASEELRALRNLFOST----- 212
QY 84 STIIGAGIAGSSEALDV-LSOAMETADPLOQLLVLSAVSGHLKTSDDLEFKALASPYY 142
Db 213 -----AGKIVSRKLEKTEETEASVRLKPGKVOVA-----IKKAVDSKYTA 252
QY 143 IRLFAVRLANLKN-TVIDIHLHSF-IHKLRPELQCSAA-IFLRLETESDAYIRDLA 199
Db 253 IPIGVFOFSCKKCPSELISHMOSIRKHOLOPDMLSKAFAVRSFL-----AFIKHLRT 304
QY 200 AKKSAISATVALOGEYQOKRFLPRLNLTLSASPOD-----OEAI 240
Db 305 AKKEET-----LQIKAKENKEVLPOLYDAVTSQOTPDSDLAIIDLDFKTESVYLQERF 359
QY 241 LYA-----LGLKRDGOSYVNIKKOLOKPDVDTPLAA----- 271
Db 360 LYACAFASHPDEELLRALISKFGSGSNDIRE-----SWMIITGALVRLKCONOGCKLK 414
QY 272 -----AQLNLAKGKEDALPIYKQALBEERRALYALRHPSELGIPALPIFLK-TKNS 326
Db 415 GVIEAKKLILIGLE-----KAKKEDIWYLLALKN-----ARLPGIPLLLKYETGE 463
QY 327 AKLWVALLELGEDPFLLEYITERLVOP-----HYNETL-----ALSFSGRT 371
Db 464 GPISHLATLQIRDVP-----FITDEVKKTMMNRIYHONRKRIHERKVTYTTAAIILKNPS 519
QY 372 LONKRVNIIVPODERERLLSTTRGEOQLITFEFLRPEAVLPCTYLLAS----- 425
Db 520 YMEVKKNILSIGELPKEMNKNYMLS---IYODILR--FETPASKVVRQVLEKEMVAHNDRF 574
QY 426 OKQOLATTAISFLSTSHOE---ALDILPQAKLPCEPIIRAYADLAIVLTLTDPKKRS 482
Db 575 SKSGSSSAYIGYVERISHSASTYSLDILYS-----GSGILRR-SNLNIFYOT---EKP 624
QY 483 LHDYAKKLIOETL-LFVDT-----ENORHPSPKPYLRV--QVTPESRTLMIDILETTLA 533
Db 625 LHGIQVIVTAQGLEALIAAPRDECEENLDYAGISALLFPVQLRPVTFNGSYDMSKML 684
QY 534 TSKSS-----EDIRLLIQLMTE-----GDAKNRPVYLAGLLIKI 566
Db 685 SASDPMSPVYKGLLLLIHDSOELQLOSGLKAMNDVQGLAIDI 727

RESULT 10
US-08-592-126-148
? Sequence 148, Application US/08592126
? Patent No. 5821091
? GENERAL INFORMATION:
? APPLICANT: Gregory Dolganov
? TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
? TITLE OF INVENTION: Polypeptides
? NUMBER OF SEQUENCES: 151
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Dehlinger & Associates
? STREET: 350 Cambridge Avenue, Suite 250
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94306
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/592,126
? FILING DATE:

```

```

? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Sholtz, Charles K.
? REGISTRATION NUMBER: 38,615
? REFERENCE/DOCKET NUMBER: 4600-0111
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 324-0880
? TELEFAX: (415) 324-0960
? INFORMATION FOR SEQ ID NO: 148:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1312 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? HYPOTHEICAL: NO
? ORIGINAL SOURCE:
? INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
? OS-08-592-126-148

```

Query Match	4.0%;	Score 112;	DB 2;	Length 1312;
Best Local Similarity	20.6%;	Pred. NO. 0.32;		
Matches 114;	Conservative 91;	Mismatches 216;	Indels 132;	Gaps 25;

01  
 02  
 03  
 04  
 05  
 06  
 07  
 08  
 09  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524

```

CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

```

```

Query Match          4.0%; Score 112; DB 2; Length 1312;
Best Local Similarity 20.6%; Pred. No. 0.32;
Matches 114; Conservative 91; Mismatches 216; Indels 132; Gaps 25;

QY 62 LKRGEDYLNKSISSPOTRKSTITIGAGLSSDALDYISQAMETADPLQOLL----- 115
DB 518 LKRLDDE-MEOLNHTTTRTOME-MLTRKDKADKDEQIRIKS--RHSDELTLGLYFPNK 573
QY 116 -----VLSAVSGHLGKTSDDL--LFKALASPYPIRLEAANRLANKTKYIDILHSFIH 168
DB 574 KQLEDWILHSSKEINQTRDLAKLAKNELAS-----SEQKNHNINELK 616
QY 169 KLPEEIOGLSAIF-----LRLTEESDAYIRDLIAAKKSATRSATLQIG 214
DB 617 RKEQLSYEDKLELDVCGSDFEESDLRLKEIEKSSKQ-RAMLAG-ATVANSQFITQLT 674
QY 215 EYQQRKRLPLRLNLTASQOD-----QEAIVLALGKLKDGQSYNNIKQLOKPDVYTL 269
DB 675 DEMOSCCPVQRYFQTAELQEVISDQSKLRLAPDKLSTSESLK-KKEKRDEMGLV 733
QY 270 AAAQALIALCKEEDALPVIK-----KQALEEPRALYARLHNPSEIGIFIAL 316
DB 734 PMQSIITDLKEK--PELNRKLQNVNRDIOFLKNDIEQSTLTGT--MPEESAKVCL 789
QY 317 P-----IFLTKNSBAKLINALLELCDDPRLLEYIT-ERLYOPHYNETLALFSK 368
DB 790 TDVTIMERFQWELKDYERKIAQQAQKLG-GIDIDRTVQVQNOEQKQKHLDITSSKIEL 848
QY 369 GRLQNMKRNVIIPQDPQR-ERLSTGTGLEQOILTPFLPKKAYL----- 416
DB 849 NRRL-----IDQDOBOIOHLKSTNELKSEKLOISTNLOROOLEQOVELSTEV 898

```

```

QY 417 PCYKLLASQKTQ---LNTAISFLSHSHQFALDLPOAKLBPSPITRAYADAIYVL 473
DB 899 OSLYREIKDAKQVSPLETTLEK-----QDEKEELINK-----NTSKKAQDKL 944
QY 474 TKDPEKRSLLHDYAKKLIQETLLFVDTENQRHPSPMYLRYGVFPESRTKMLDIETLA 533
DB 945 NDIKREKVNINHYMK-----DIENYIDQGDVDYRKQKQETELNKYIADLSECEK-H 993
QY 534 TKSSEDIRLLIQ 546
DB 994 KEKINEDMKLRMO 1006

```

```

RESULT 12
US-08-466-390-4
Sequence 4, Application US/08466390
Patent No. 5686562
GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,390
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

```

```

Query Match          3.98%; Score 111; DB 1; Length 2101;
Best Local Similarity 21.9%; Pred. No. 0.85;
Matches 113; Conservative 81; Mismatches 184; Indels 138; Gaps 25;

QY 38 QSTQALATYLEALDAYGDHDFVLRKIGEDY-LKQSIHSSDPQTRKSTIIGAGLASSSE 96
DB 1387 QSKQAAGGLARELLRAQ-----RELGELIPKQVYAE---QERTAQOIRAKKASYAE 1435
QY 97 ALDVLASQAMETADPLQOLLVLSAVSGHLGKTS-----DLFKALASPYPIVIRLE 146
DB 1436 QLSMLKKRAHG-----LLAEENRGLGERANRGQFLEVEDQAREKQYQELAAVRAD 1486
QY 147 AAYRLAMUK-----NTKYVDHLHSPTIHLKLPEDIOCSAIFRLRLE 186
DB 1487 AETRLAEVQREAOASTARELEVMTAKYEGAVKYLEERQRF---QEBRQKLTIA-----Q 1536
QY 187 TEESDAYIRDLIAAKKSATRSATALQI--GEYQO--KRFLPTLNLNLTASPODEAIVL 242

```

Db	1537	VEBSSKLLADSDASVQOQKLUKVAQAGSGESQEQEAGRRQQAQNLBQAQNSQKBEAAEHY	1596
OY	243	ALGKIKGOSYVYIKK-----QIQPDVDTYLAQAALIALGKEDALPYI	288
Db	1597	KL-OMEKAKHHYKAKOONOELOEOLRLEOLOKENEKELRAEAE---LGHLELOAQGLK	1651
OY	289	KQALTEERPALVA-LRHLPSFEG-----IPILPFLFKTNSEAK--LNVALA	334
Db	1652	TKEA-EOTCHNLTAQVRSLEAQVAHADQOLRDLGKFOVATDA-LKSRBPQKPOLDSID	1709
OY	335	LLEIGCDTPKLELYTERLYOVPHYNETLALSFSGKGTIONMKRVNIIVPDQOE--REKL	392
Db	1710	SLDLSCEEGTPLS-ITSKLPTROPDPT-----SYGEGPASPISQRL	1749
OY	393	LSTTRGEEQILFLF--FLRPEAVLPCTIKLL--ASQQTOLATTAISPLSHSBOAL	447
Db	1750	PKWESLESILYFPIIPARQAPLESSLDSLGVPFLDSGKTKTSARRRTOIINITMKLL	1809
OY	448	DLEFOAKKLGEPEPIRAYADLAITYNLTKDEKKRSI	483
Db	1810	DV-----EEP-----DSANSSFYSTRSAPASASL	1834

Query Match	3.9%	Score 111;	DB 1;	Length 2101;
Best Local Similarity	21.9%	Pred. No. 0.85;		
Matches 113; Conservative	81;	Mismatches 184;	Indels 138;	Gaps 25

OY	38	OSTOOALTYLEALDAYGDHDFFLKRGEDY--LKOSIHSSPDQTKSTITIGAGLASSE	96
Db	1387	OSKQAAGLRARELLRAO-----RELIGELPIRLKVAE--QERTAOOLAREKRAVSVE	1435
OY	97	ALDVLISQMETADELQOOLLVLSAVSGHIGKTSD-----DLFKALASPPYVIRLE	146
Db	1436	QLSMKKAHG-----LLAEENRGLGBRANMLGQFLEVELDQAREKXYOVELAANVRD	1486
OY	147	AAVRLANK-----NKTVIDHASFTHKLPREIOCLSAAIFLRLE	186
Db	1487	AETRLAEYORASTANELEVMYAKYEGAKVKLEEROR-----OBEROKFLA-----Q	1536
OY	187	TEESDAYIRDLIAAKKSAIRSATALOI--GEYOO--KREPTLRNLTASPODEALYI	242
Db	1537	VEELSKRLADSDQASKVQOQKIKAAVOAGGSESOEQAORFOAQUNELQANOLSKOEBAEHY	1596
OY	243	ALGLKXOCOSYNNKK-----QLOKPDVDTYLAQAQALLMGKEEDALPYI	288
Db	1557	KL--OMERAKTYTDYDKKCONGELOEOLRSLBEOLOKEMKEELRAEABER---LGHELOAGLK	1651
OY	289	KKQALEERPALYA--LHRLSEIG-----PIALPIFLTKNSEAK--LNAVLA	334
Db	1652	TKEA--EOTCRHLLTQAVRSLAEQVAHMDQOLRDLGRFOVATDA--LKSREQAORFOLDSTD	1709
OY	335	LLEIGCOTPKLLEIETIRLVQPHNYNLTALSPFSKGTLONMKRVNIYPODQOF--REPL	392
Db	1710	SLDLSCEEGPPLS--TTSKLRTPDDGT-----SVYGEASPISORL	1749
OY	393	LSTTRGLSEOLITFL--FRPKAEVILPCIIYKLL--ASQGTOLATTAISFLSHTHOBAL	447
Db	1750	PKPVESLESILYFPIPARSQAPLESLSJGDFLDSGKTKTSARRRTOIINTMYTKL	1809
OY	448	DLFLQAALKECEPTIRAYADLAIYNLTKPOEKKRSL	483
Db	1810	DY-----EEP-----DSANSSFYSTRSAPASQOST	1834

RESULT 15  
 US-08-483-924-4  
 : Sequence 4, Application US/08483924  
 : Patent No. 3882876  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: TOKKATLY, GARY  
 : APPLICANT: LIDGARD, GRAHAM P  
 : TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE  
 : TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX  
 : NUMBER OF SEQUENCES: 6  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
 : STREET: 125 HIGH STREET  
 : CITY: BOSTON  
 : STATE: MA  
 :  
 : COUNTRY: USA  
 : ZIP: 02110  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/483, 924  
 : FILING DATE: 07-JUN-1995  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: PITCHER ESO, EDMUND R  
 : REGISTRATION NUMBER: 27,829  
 : REFERENCE/DOCKET NUMBER: MTP-013

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: (617) 248-7000
? TELEFAX: (617) 248-7100
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2101 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-08-483-924-4

```

Query Match Similarity 3.9%: Score 111; DB 2; Length 2101;  
Best Local Similarity 21.9%: Pred. No. 0.85; Mismatches 184; Indels 138; Gaps 25;  
Matches 113; Conservative 81;

QY 38 OSTQOALATYTLBALDAYGDHDFVLRKIGEDY-LKQSHSDPDQRTKSTTIAGIAGSSE 96  
DB 1387 OSKQOAGGLRABELLRNQ-----RELGEILPLNQKAE--QERTMQQLRAEKASVAE 1435  
QY 97 ALDVSQAMETADPLQOQLLVSAVSGHGLKTSQ-----DILFQALASPYVIRLE 146  
DB 1436 QLSMKKKHAG-----LLAENRGIGERANLRGOLFVELDQAREKYVQELAAVRAD 1486  
QY 147 AAYRLANK-----NFWIDHLSFIHKLPDEIQCSAALFLRE 186  
DB 1487 AETRLAEVQREKQSTARELEVMTAAVEGAKKAVLEEROF---OEERKKTLA-----Q 1536  
QY 187 TEESDAYTRDLLAAKSAIRSATALQI-GEYQO--KREPLTRLNLTSASPODEAILY 242  
DB 1537 VEELSKTKLADSDQASKVQOQK LKAVQAOGEGESQQAORQADUNELQALQSGKEGAEHY 1596  
QY 243 ALGKLKDGQSYNNIK-----QIQKRDVDTYTLAAQALALAKGEEDALVYI 288  
DB 1597 KL-QWEKKATHYDARKKQNOQLQELQSLSEQLQKENEKELRAEAE---LGHLEQOAGLK 1651  
QY 289 KQALAEERPRLAY-LRHLSPSIG-----IPLALPIFLTKNSSEAK--LUNVALA 334  
DB 1652 TKEA-EQYCRHNLTAQVRSLENAQVAADQDLRGKFGVAYTDA-LKSREPOAKPQDLSDID 1709  
QY 335 LLELGGDTPKLLVETRYTERLVOPHYNETLALSFGKRTLQNMKRVNIIVPODQE--RERL 392  
DB 1710 SLDSLCEGTEPLS-ITSKLPTPOPGT-----SVPGESPASPISQRL 1749  
QY 393 LSTTNGLEQLITFL---FLPLKAEVLEPCYIKLL--ASQKTQALTAITSFSHTSHQAL 447  
DB 1750 PPKEYESTSLYFTPIPARSQAPLBSLSDSLDVFELDSGRKTRRSARRRTTQIINTMTKKL 1809  
QY 448 DLLEQAAKLPEPIIRAYADLAIVNLTKDPEKKSL 483  
DB 1810 DY-----EFP---DSANSEFSTRAPSPQASL 1834

Search completed: August 20, 2002, 04:11:08  
Job time: 3293 sec